C.1.

TOWARDS A CONSENSUS OF OPINION

by

CHRISTIAN GENEST

B.Sp.Sc., Université du Québec à Chicoutimi, 1977
M.Sc., Université de Montréal, 1978

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Mathematics

We accept this thesis as conforming

to the required standards

THE UNIVERSITY OF BRITISH COLUMBIA

January 1983

Department of Mathematics

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date: January 25, 1983

Abstract

This thesis addresses the problem of combining the prior density functions, $f_1,\dots,f_n$, of n individuals. In the first of two parts, various systems of axioms are developed which characterize successively the linear opinion pool, $A(f_1,\dots,f_n) = \sum_{i=1}^{n} w_i f_i$, and the logarithmic opinion pool, $G(f_1,\dots,f_n) = \prod_{i=1}^{n} f_i^{\alpha(i)} / \int \prod_{i=1}^{n} f_i^{\alpha(i)} d\mu$. It is first shown that A is the only pooling operator, $T(f_1,\dots,f_n)$, which is expressible as $T(f_1,\dots,f_n)(\theta) = H(f_1(\theta),\dots,f_n(\theta),\theta)$ for some function H which is continuous in its first n variables and satisfies $H(0,\dots,0,\theta) = 0$ for μ- almost all θ. The regularity condition on H may be dispensed with if H does not depend on θ. This result leads to an impossibility theorem involving Madansky's axiom of External Bayesianity. Other consequences of this axiom of group rationality are also examined in some detail and yield a characterization of G as the only Externally Bayesian pooling operator of the form $T(f_1,\dots,f_n)(\theta) = H(f_1(\theta),\dots,f_n(\theta))/\int H(f_1,\dots,f_n)d\mu$ for some $H:(0,\infty)^n \longrightarrow (0,\infty)$. To prove this result, it is necessary to introduce a "richness" condition on the underlying space of events, (Θ,μ). Next, each opinion f_i is regarded as containing some "information" about Θ and we look for a pooling operator whose expected information content is a maximum. The operator so obtained depends on the definition which is chosen; for example, Kullback-Leibler's definition entails the linear opinion pool, A.

In the second part of the dissertation, it is argued that the domain of pooling operators should extend beyond densities. The notion of propensity function is introduced and examples are given which motivate this generalization; these include the well-known problem of combining P-values. A theorem of Aczél is adapted to derive a large class of pooling formulas which encompasses both A and G. A final characterization of G is given via the interpretation of betting odds, and the parallel between our approach and Nash's solution to the "bargaining problem" is discussed.

James V. Zidek

Thesis supervisor

Table of Contents

List of Figures

Technical Note

This thesis was prepared on the Amdahl 470 V8 computer of the University of British Columbia with the aid of the FMT text-processing language. Because the character sets for the Xerox 9700 printer are somewhat limited, it was necessary to depart slightly from some conventional mathematical symbolism. For instance, the letter "R" had to be reserved to denote the real line and the British Pound Sterling symbol "£" was substituted for script ell. On some occasions, it was also necessary to write subscripts on the same line as the indexed quantities, e.g. w(i) instead of w_i when this quantity appeared as an exponent. Furthermore, tildas were systematically printed over the variables instead of under. We hope the reader will not be inconvenienced by these departures from common usage.

The material is divided into 4 chapters, and each chapter into several sections. Equations, definitions, theorems and examples are numbered in the decimal notation. Thus, Equation (2.3.5) refers to the fifth labelled equation of Section 3, in Chapter 2. Within Section 2.3, it might be referred to simply as Equation (5).

Acknowledgements

The author wishes to thank his supervisor, Prof. James V. Zidek, for suggesting most of the problems treated in this thesis and for financial assistance. His appreciation is also extended to Mr. Bruce J. Sharpe for his constant willingness to enter into discussion.

The financial support of the National Science and Engineering Research Council and of the University of British Columbia are gratefully acknowledged.

Dedication

This dissertation is dedicated with all my love to the two women who have supported me throughout my studies: my mother, Lucie Lapointe-Genest, and my wife, Christine Simard-Genest.

I. PROLEGOMENA

1.1 Introduction

In this thesis, we shall be concerned with the problem of devising methods for aggregating opinions. By "opinion," we mean the expression of a person's belief vis-à-vis the outcome of an uncertain event, as opposed to an intention, as in "opinion polls" (e.g., "whom do you intend to vote for?").

Usually, opinions will be expressed as (subjective) probability distributions over the appropriate space Θ of "states of nature." They might be prior, posterior, structural (Fraser 1966) or fiducial distributions, for instance. Indeed, provided that the "degrees of belief" of an individual are assessed quantitatively and in a coherent manner, they can be shown to conform to the axioms of probability theory (de Finetti 1937; this question has recently been reexamined by Lindley 1982).

However, common observation and experimental studies (Winkler 1967; Tversky & Kahneman 1974; Slovic et al. 1977) tend to confirm that although an individual may have a good knowledge of the relative likelihood of the various possible states of nature, it cannot generally be expected that he will also master the calculus of probabilities and express his opinion accordingly. Moreover, we would like to account for the use of such widely spread expressions of belief as improper, vague, and uniform or non-informative priors, as well as the more recent concept of belief function discussed by Shafer (1976).

Therefore, we shall take an opinion to be any function $f:\Theta \longrightarrow [0,\infty)$ on the space Θ of possible states of nature. If the range of f is restricted to $(0,\infty)$, we will call it a propensity function, or P-function for short. Furthermore, if f belongs to a collection $\Delta = \{g:\Theta \longrightarrow [0,\infty) \mid \int g d\mu = 1\}$ of probability densities with respect to a dominating measure μ on Θ, we will say that f is a μ-density.

In accordance with the literature, we will call assessor or expert any person who is asked to produce his opinion concerning Θ. Generally, an assessor will be some kind of expert whose opinion on the subject matter is deemed to be enlightening, but it need not be so. In our discussion, it will always be assumed that when asked for their opinion, the experts are capable and willing to present an assessment of all relevant facts and evidence known to them. As long as this is the case, every opinion has its value and should be treated with consideration. For, in subjective or personalistic probability theory, the relative likelihood attributed to an event or hypothesis is simply what the assessor believes it to be; there is no such thing as a "correct" or "objective" opinion. For a critical review of this approach, cf. Fine (1973).

1.2 The problem of the panel of experts

In a decision analysis, it is often necessary to combine a group of individuals' beliefs into a single representative opinion which may be thought of as a consensus of these peoples' judgements. This problem of determining decision rules for statistically aggregating individual opinions without group discussion nor bargaining is called the problem of the panel of experts, after Raiffa (1968).

Let us suppose, for instance, that a decision maker who has very little knowledge of a subject-matter is confronted with the need to quantify his beliefs and determine a prior distribution before he can undertake a formal Bayesian analysis. To solve this problem, he might choose to express his ignorance by using a prior distribution which adds little to the sample information; much work has been done along these lines by Lindley (1961), Jeffreys (1967), Novick & Hall (1965) and Zellner (1977). However, the uncritical use of non-informative prior measures sometimes leads to inconsistency and paradoxes such as those presented by Dawid, Stone & Zidek (1973) or Stone (1976). Moreover, this approach will not be satisfactory if the problem at hand is of major importance and its analysis will result in a costly, irreversible decision. If time is pressing or collection of a large amount of data is impractical, the decision maker may well decide to consult one or several people who do have knowledge believed to be relevant to the question, i.e. experts. These experts are said to form a <u>panel</u> of consultants. Even after extensive discussion amongst them about their beliefs and proper modification of their respective opinions to take into account all the jointly perceived and available information, it is unlikely that the experts will converge to a state of total agreement. When this happens, we say that the group is left in <u>dissensus</u>. As only <u>one</u> opinion is needed in the end, how does the decision maker proceed to extract it from the number of (possibly) diverging opinions he has collected, without being irresponsive to any particular assessment?

Despite the prevalence of consulting in countless decision-making situations, this problem has received comparatively little attention in the literature, as pointed out in the review papers of Winkler (1968) and Hogarth (1975,1977). An instructive introduction to the various theoretical questions raised by group-assessments is provided by Raiffa (1968).

In conceptualizing the phenomenon, we will assume that all discussion and argument have taken place at the time when the decision maker is presented with a number, n, of opinions, one for each member of the panel of experts. We will place ourselves in his shoes and attempt to find desirable properties which a consensus opinion should have. More specifically, we intend to propose and explore the consequences of various interpretations of such vague concepts as "adequacy," "representativeness" and "consensus."

The approach we take here is normative, in the sense that we prescribe the way in which a decision maker should process expert opinions if he wishes to adhere to certain postulates of coherence and rationality. No attempt is made to describe how a person confronted with the actual problem would be observed to carry out the task. Moreover, we are implicitly adopting the view that in the presence of uncertainty about Θ, the quantity of interest to the decision maker is $T(f_1,\ldots,f_n)$, the consensus opinion describing his final assessment of beliefs concerning the possible outcomes of the event upon collecting the experts' views $f_1,\ldots,f_n$.

This attitude is by no means generally accepted, although it has some supporters (Winkler 1968; Weerahandi & Zidek 1978; Bernardo 1979 in a different context). However, it makes perfectly clear that the problem we propose to examine differs markedly from that of a group faced with a decision-making situation, where the main concern lies with the final group decision and where -by way of necessity- considerations of utility and bargaining are invoked. The distinctions and relations between these two problems have been well emphasized in a survey paper by Weerahandi & Zidek (1981).

Admittedly, therefore, the expression "decision maker" is something of a misnomer. In fact, the present set-up leaves open the possibility that the decision maker represents the "synthetic personality" of a group amongst whose members discussion has failed to create a consensus, but which is called upon, nevertheless, to produce a single assessment of beliefs representative of the various opinions voiced. This would be the case if, for example, a group of meteorologists was asked to issue a joint forecast.

It is important to realize, then, that the role played by our decision maker has a lot in common with that of the statistician of standard decision theory when he attempts to devise strategies to estimate a quantity from a number of observations. Only here, observations are opinions, the estimate sought is also an opinion, and there is no "true objective opinion" against which to judge the performance of the pooling formula that the decision maker chooses to use.

1.3 Previous proposals

The literature which relates to the so-called problem of the panel of experts divides roughly into two main streams: on the one hand are those papers which deal directly with personal probability distributions and their aggregation; on the other are those which are concerned with a broader picture, that of a group decision problem, and which focus on considerations of utility rather than on the consensus of opinions. References of the latter type will only be mentioned when they include some helpful comments regarding the problem at hand.

Following Stone (1961), most investigators have represented group assessments by taking a weighted average of individual distributions. Formally, if f_i represents the probability density of the i-th member of the group concerning the unknown quantity Θ, then the linear opinion pool for n experts is defined by

$$T(f_1,\ldots,f_n) = \sum_{i=1}^{n} w_i f_i , \tag{1.3.1}$$

where $w_i \geq 0$ and $\sum_{i=1}^{n} w_i = 1$. The restrictions on the weights, w_i, insure that the joint opinion of the group, $T(f_1,\ldots,f_n)$, will be a density.

Winkler (1968) discusses the problem of determining the weights for each expert's distribution and proposes various ad hoc solution schemes. In subsequent studies, he and others (Staël Von Holstein 1970; Winkler 1971) found that the uniform weighting scheme, $w_i = 1/n$, was never outperformed more than marginally (in terms of predictive ability) by other schemes which attempted to rank each expert according to his expertise or past performance (these included the method of Roberts (1965) for combining and updating weights using Bayes' Theorem). Also supporting Formula (1.3.1) is the often observed empirical fact that composite distributions show greater predictive ability than most of the individual experts, [1] a phenomenon which might be linked to the reliability of average point estimates in the classical theory of estimation.

But it was in the work of Bacharach (1973, 1975) that solid theoretical grounds for implementing the linear opinion pool began to emerge. Although Bacharach is mainly concerned with the existence of a sensible group preference relation for ordering the possible courses of action in the face of differences of opinion <u>and</u> utility, he finds conditions on this group preference relation which entail, as a unique solution, Stone's Formula (1.3.1). In a theorem which he attributes to Madansky (1964), Bacharach derives (1.3.1) by assuming essentially that the group will prefer an action A to another action B whenever each of its members does, and that the group's preference relation is not affected by the presence of irrelevant alternatives. After arguing in favour of these

---

[1] Dr. F. P. Glick brought to my attention some recent work of Alan Shapiro (1977,1979) who rediscovered this fact for himself and used the linear opinion pool to increase diagnostic accuracy of physician-experts.

postulates, he goes on to show that if an extra condition which he calls group rationality is introduced, the pool can be forced into dictatorial form, i.e. one of the w_i's of (1.3.1) equals 1 while the remainder are 0. This far-reaching condition asserts that the group acts as if it were a single expected-utility maximizer. In Section 2.4, we too will prove such an "impossibility theorem" using a similar concept called by Madansky (1964,1978) External Bayesianity.

In our endeavour, we have been very much stimulated by the work of Kevin McConway (1981) who was the first to give a strong justification for using the linear opinion pool within the framework we propose to adopt ourselves. His main theorem states that if a decision maker wants his process of consensus finding to commute with the marginalization of the distributions involved, then he has no alternative but to use Formula (1.3.1). Chapter 2 of this thesis will start with a discussion of McConway's result.

Despite its great popularity, the weighted average formula is endowed with features which may in certain circumstances be viewed as drawbacks. For instance, Winkler (1968) notes that (1.3.1) is typically multi-modal on its domain and so may fail to identify a parameter which typifies the individual choices. This leads him to formulate an alternate prescription, which he calls the natural conjugate approach. In this method, each group member's opinion is deemed to constitute "sample evidence" which can be represented by a natural conjugate prior (Bickel & Doksum 1977, p. 77) to the distribution of the data-generating process of interest. In order to form the group assessment, therefore, the decision maker need only combine opinions in a manner similar to successive applications of Bayes' Theorem. The reader will recognize that even if this approach obviates the possibility of multi-modal distributions, it leaves us nonetheless with the difficult question of determining weights for each of the experts, as well as the degree to which their opinions are based on overlapping experience and data sources.

It is precisely these difficulties which brought Morris (1974,1977) to elaborate a theory of expert use that is entirely consistent with the Bayesian philosophy. In his work, Morris pushes Winkler's idea one step further and treats each expert probability distribution as a <u>random variable</u> whose value is revealed to the decision maker. To obtain the consensus distribution, the decision maker must then proceed to introspect a likelihood function representing his assessment of the different experts' knowledge and combine their opinions one-by-one with his own using Bayes' rule. A conventional, albeit complex uni-Bayesian analysis results. In the simplest case, that where the members of the panel rely on independent data bases and are good probability assessors (i.e. they are calibrated[1]), Morris shows that the normalized product of the experts' priors obtains:

$$T(f_1,\dots,f_n) = \prod_{i=1}^{n} f_i^{\alpha(i)} \Big/ \int \prod_{i=1}^{n} f_i^{\alpha(i)} \, d\mu, \tag{1.3.2}$$

with $\alpha(i) = 1, 1 \le i \le n$; this is a particular case of the so-called logarithmic opinion pool which Dalkey (1975) had earlier proposed on an ad hoc basis. Note that for $T(f_1,\dots,f_n)$ to be a μ-density in (1.3.2), the $\alpha(i)$'s do not need to add up to one. However, Morris assumes that the decision maker is one of the members of the panel, which forces $\alpha(1) = \dots = \alpha(n)$; otherwise, the solution would be expert-dependent and thus could not constitute an acceptable consensus.

Although Morris' work certainly is conceptually appealing, consistent and insightful, only few would agree with him that it represents a practical methodology because of the insurmountable assessment problems which overlapping experience would cause. For this reason, other researchers have continued to seek direct ways of pooling opinions.

It is interesting to observe that the prescription embodied

---

[1] As Lindley & al. (1979) observed, this is asking a lot. On the other hand, Dawid (1982) has recently shown that a coherent Bayesian expects to be well calibrated!!

in (1.3.2) is not only a natural implication of Morris' analysis but as well it is Externally Bayesian when the $\alpha(i)$'s add up to one. This axiom, formulated by Madansky (1964,1978), requires the commutativity of the operations of (i) compounding individual probabilities into a group probability; and (ii) updating probability assessments via Bayes' formula. The basic properties of Externally Bayesian procedures have been explored by Madansky (1978) who finds that of those pooling formulas which are discussed in the literature, only dictatorships, (1.3.2) and certain applications of the natural conjugate approach survive the test.

External Bayesianity has also been supported by Weerahandi & Zidek (1978), who call it prior-to-posterior coherence. Their unpublished manuscript contains an incorrect (but correctable, cf. Theorem 2.4.6) derivation of the logarithmic opinion pool using External Bayesianity. Moreover, they argue that randomized decision rules need to be introduced because, in some cases, the prior opinions of the experts are so discrepant that "tossing a coin" is the only satisfactory means for choosing between them. In Section 2.5, a similar idea will lead us to still another characterization of Stone's linear opinion pool, whilst it will be shown in Section 2.4 that the logarithmic opinion pool is, in some sense, the only Externally Bayesian procedure for combining the prior density functions $f_1,\ldots,f_n$ of n individuals.

Our survey of the literature on the problem of pooling densities would not be complete without at least a brief mention of the two following papers.

In 1959, Eisenberg & Gale (see also Norvig 1967) presented an ingenious scheme for combining probability distributions based on the "pari-mutuel" betting method. Their idea was that the principles determining "totalisator odds" could appropriately determine group judgements in more general situations. However, it can be shown that certain individual opinions will allow their holders to dictate the consensus odds, and for that reason, the pari-mutuel method has never been very popular.

Then, DeGroot (1974) proposes that upon being apprised of the distributions of the other members of the panel, each expert updates his own prior using Formula (1.3.1) by assigning importance weights to himself and his peers. The procedure is then iterated until further revisions no longer alter any of the members' opinions, and limit theorems from Markov Chain theory are invoked to determine when a consensus distribution exists and what it is when it does exist. Berger (1981) points out an error in DeGroot's original paper and gives the exact conditions under which the iterative process will converge. A variation on the theme is described by Press (1978), who also provides an extensive list of references.

Although DeGroot's method does not formally fit our set-up (remember that we assume all discussion is closed), it could be imagined that the iterative process is carried out by the decision maker himself, once he has taken note of the various opinions expressed as well as the ratings that each individual expert granted to himself and to the other people who were consulted. Along with French (1981), however, we find three important flaws in this general approach. They are: (i) the linear opinion pool is still proposed as an ad hoc procedure; (ii) it is assumed that no outside data, observations or information about the value of θ is available (this invalidates External Bayesianity as a selection principle); and last but not least, (iii) no provision is made for the (non-negligible) case when the iterative procedure leaves the group in dissensus.

1.4 Outline of subsequent chapters

As pointed out in the first section of this chapter, we do not restrict ourselves to what could be called the classical form of the problem of the panel of experts, where opinions are assumed to have been expressed as probability distributions over Θ. Rather, we enlarge the definition of opinion to include <u>any</u> expression of beliefs, $f:\Theta \longrightarrow [0,\infty)$, integrable or not. Thus, we will be concerned with propensity functions as well as density functions. The two topics will be discussed in separate chapters.

Starting with the problem of pooling densities, Chapter 2 begins with a discussion of McConway's (1981) characterization of the linear opinion pool via the Marginalization Property. A short proof of his theorem appears in Section 2.2. We are then led to propose in Section 2.3 a different derivation of the same operator founded on the concept of "locality." At least one avenue for generalization is explored. Section 2.4 is devoted to a study of Madansky's (1964;1978) axiom of External Bayesianity and the logarithmic opinion pool is shown to be the only quasi-local pooling formula which is consistent with this postulate. Some of the general properties possessed by Externally Bayesian procedures are derived on the way. In Section 2.5, we regard a consensus distribution as a statistic which condenses the information contained in a set of opinions. The idea of an information maximizing pooling operator then leads to yet another characterization of Stone's linear opinion pool. Furthermore, a large class of pooling formulas containing both the linear and the logarithmic pools as limiting cases is derived using Kullback's (1968) notion of divergence. Finally, Section 2.6 discusses these findings and reiterates some words of caution.

Chapter 3 addresses the more general problem of combining a number of propensity functions. Examples are presented in Section 3.1 which motivate this generalization; amongst them appears the well-known problem of combining independent tests of hypothesis. All these examples are continued in Section 3.2, where we focus our attention on pooling operators which are local. Here, we argue that the quasi-arithmetic weighted means of Hardy, Littlewood & Pólya (1934) are the only "sensible" local rules for aggregating propensities; the approach is axiomatic and well suited to those cases where the experts' scales of belief are intercomparable. The last section is devoted to characterizing the logarithmic pool when the comparability assumption is not met; the parallel between our approach and Nash's (1950) solution to the bargaining problem is also discussed.

Finally, Chapter 4 contains suggestions for further research.

II. POOLING DENSITIES

2.1 Fundamentals and notation

Throughout this chapter, we will denote by Θ the space of (mutually exclusive) states of nature over which each of a group of n assessors ($n \geq 2$) will be asked to produce a probability distribution. For convenience, we will assume that μ is a dominating measure on Θ and that each opinion is expressed as a μ-measurable function $f:\Theta \longrightarrow [0,\infty)$ with $\int f d\mu = 1$. Because of the properties of the Carathéodory process for generating measures, there is no loss of generality in assuming μ to be complete and in taking $\Omega(\mu) = \{A \subset \Theta | \mu(B)=\mu(B \cap A)+\mu(B \setminus A)$ for all $B \subset \Theta\}$ to be the σ-field of μ-measurable sets. In applications, μ will usually be σ-finite, so that every other measure ν which is absolutely continuous with respect to μ will have a Radón-Nikodym derivative (Sion 1968, p. 110). But σ-finiteness is not essential otherwise.

We will write Δ for the collection of all μ-measurable densities on Θ, and $(f_1,\ldots,f_n)$ to represent either a typical element of Δ^n or else the function $f:\Theta \longrightarrow [0,\infty)^n$ defined by $f(\theta) = (f_1(\theta),\ldots,f_n(\theta))$. The interpretation will always be clear from the context, so no confusion will arise from this convention. It will also be assumed that $\Delta \neq \emptyset$, so that there exists $A \in \Omega(\mu)$ with $0 < \mu(A) < \infty$.

By a <u>pooling operator on</u> Θ, we mean any application $T:\Delta^n \longrightarrow \Delta$ which maps the n-tuple $(f_1,\ldots,f_n)$ to $T(f_1,\ldots,f_n)$, a μ-density. The following definition lists the properties of pooling operators which we will most often refer to.

<u>Definition 2.1.1</u>

We say that a pooling operator $T:\Delta^n \longrightarrow \Delta$ is

(1) <u>local</u>

iff there exists a Lebesgue-measurable function $G:[0,\infty)^n \longrightarrow [0,\infty)$ such that $T(f_1,\ldots,f_n) = G\square(f_1,\ldots,f_n)$ μ-a.e. Here, $\square$ represents the usual composition of functions.

(2) <u>quasi-local</u>

iff there exists a function $C:\Delta^n \longrightarrow (0,\infty)$ such that $T(f_1,\ldots,f_n)\cdot C(f_1,\ldots,f_n)$ is local.

(3) <u>unanimity preserving</u>

iff $T(f_1,\ldots,f_n) = f$ μ-a.e. whenever $f_i = f$ μ-a.e. for all $1 \le i \le n$.

(4) <u>dogma preserving</u>

iff $\text{Supp}(T(f_1,\ldots,f_n)) \subset \bigcup_{i=1}^{n} \text{Supp}(f_i)$ μ-a.e., where in general $\text{Supp}(f) = \{\theta\in\Theta | f(\theta) \ne 0\}$.

(5) a <u>dictatorship</u>

iff there exists $1 \le i \le n$ such that

$T(f_1,\ldots,f_n) = f_i$ μ-a.e. for all choices of $f_1,\ldots,f_n$ in the domain of T.

To prove theorems, we will often make use of the following elementary results from the <u>Theory of functional equations</u>. Their proofs are to be found in Aczél (1966).

<u>Lemma 2.1.2</u>

Let $h: R^n \longrightarrow R$ be Lebesgue-measurable or non-decreasing in each of its n variables. If

$$h(\tilde{x})+h(\tilde{y}) = h(\tilde{x}+\tilde{y}) \qquad (2.1.1)$$

for all $\tilde{x}$ and $\tilde{y}$, vectors of real numbers, then there exist constants $c_1,\ldots,c_n \in R$ such that $h(\tilde{x}) = \sum_{i=1}^{n} c_i x_i$ over its domain, $\tilde{x} = (x_1,\ldots,x_n)$. The result also holds true when the domain of h is $[0,\infty)^n$ or $[0,K]$ with $K > 0$ a constant.

<u>Lemma 2.1.3</u>

Let $h:(0,\infty)^n \longrightarrow R$ be Lebesgue-measurable or non-decreasing in each of its n variables. If

$$h(\tilde{x})\cdot h(\tilde{y}) = h(\tilde{x}\cdot\tilde{y}) \qquad (2.1.2)$$

for all $\tilde{x},\tilde{y} > \tilde{o}$, then $h(\tilde{x}) = \prod_{i=1}^{n} x_i^{c(i)}$ for some $c(i) \in R$.

Equations (2.1.1) and (2.1.2) are usually referred to as Cauchy's functional equation. In the present context, addition and multiplication of vectors is taken to be componentwise.

2.2 McConway's work in review

In this section, McConway's (1981) derivation of the linear opinion pool will be discussed. A short proof of his theorem (labelled 2.2.4) will also be offered. The discussion will motivate and serve as a background for our own results.

It has already been mentioned that the approach adopted by McConway fits the description of the problem of the panel of experts set out in Section 2 of Chapter 1. To justify the prescription embodied in (1.3.1), McConway first introduces what could be called the Marginalization Postulate (MP). This condition stipulates that the same consensus distribution should be arrived at whether (i) the assessors' distributions are first combined and then some marginalization is performed on the consensus; or (ii) each assessor individually performs the marginalization and the resulting marginal distributions are pooled into a consensus distribution.

If a mild condition tantamount to that in our definition of a dogma preserving pooling operator is added, then McConway proves that the Marginalization Postulate is equivalent to what he calls the <u>Strong Setwise Function Property</u> (SSFP), which in turn holds true if and only if the pooling operator is of the form (1.3.1). (To carry out his program, McConway introduces the "Weak Setwise Function Property" (WSFP) and proceeds to show that this property is equivalent to his Marginalization Postulate. This fact constitutes Theorem 3.1 of his paper. However, the proof is obscured by a misprint. The third line of the last paragraph, p. 411, should read "any $S \in \Sigma$ which contains A has $\sigma(A)$ as a sub-σ-algebra..." and not "any $S \in \Sigma$ contains A, has $\sigma(A)$ as a sub-σ-algebra...").

In his treatment, McConway does not assume the existence of a dominating measure μ, and consequently his results are stated in terms of probability measures as opposed to densities. However, McConway does say that "in practice, the experts will usually agree on some obvious σ-field over Θ," and we claim that most of the time, a natural μ will impose itself just as well, so that little is lost by assuming its existence. This point of view is not entirely new, as an inspection of the set-up in Stone (1961) or Weerahandi & Zidek (1978) will confirm. (Note that for μ to exist, it suffices for the experts to agree on some natural σ-additive function τ on a ring H. The Carathéodory process will then automatically extend τ to μ and H will be contained in $\Omega(\mu)$. Cf. Sion 1968, p. 67).

When a natural dominating measure μ on Θ exists, McConway's SSFP condition can be formulated as follows:

<u>Definition 2.2.1</u> (McConway 1981)

A pooling operator $T:\Delta^n \longrightarrow \Delta$ has the <u>Strong Setwise Function Property</u> (SSFP) iff there exists a function $F:[0,1]^n \longrightarrow [0,1]$ such that

$$\int I(A)\cdot T(f_1,\ldots,f_n)d\mu = F[\int I(A)f_1 d\mu,\ldots,\int I(A)f_n d\mu] \qquad (2.2.1)$$

for all $A \in \Omega(\mu)$, where in general $I(A)$ stands for the indicator function of the set A.

Before we state McConway's Theorem, we make a definition of our own: [1]

<u>Definition 2.2.2</u>

We say that a space (Θ,μ) is <u>tangible</u> iff there exist (at least) three μ-measurable neighbourhoods A_1,A_2,A_3 in Θ such that

(I) $0 < \mu(A_i) < \infty$, $i=1,2,3$;

---

[1] Shortly after the completion of this thesis, a paper of Carl Wagner (1982) was brought to our attention in which the author uses the term "tertiary space" to denote what we call a tangible space. Theorem 7 of his paper is equivalent to the formulation of McConway's Theorem presented in Theorem 2.2.4 below.

and

(II) $\mu(A_i \cap A_j) > 0 \Rightarrow i=j$.

One might wonder what an "intangible" space looks like. A more or less complete answer to this question is provided by the following lemma.

Lemma 2.2.3

The spaces $\Theta_1 = \{\theta_1\}$, and $\Theta_2 = \{\theta_1,\theta_2\}$ with ν = counting measure typify the class of intangible spaces.

Proof: Let (Θ,μ) be intangible, so that there are at most two μ-measurable sets A_1, A_2 with properties (I) and (II). Assuming $\Delta \neq \emptyset$ is enough to guarantee the existence of at least one such set A_1.

If there exists only one set A_1 satisfying (I) & (II), then the function $f = I(A_1)/\mu(A_1)$ is the only element of Δ and (Θ,μ) is clearly equivalent to (Θ_1,ν).

On the other hand, if there are exactly two sets A_1, A_2 with properties (I) and (II), then it follows from Definition 2.2.2 that any μ-measurable subset B of A_i must obey $\mu(B) = 0$ or $\mu(B) = \mu(A_i)$, $i=1,2,3$, where $A_3 = \Theta\setminus(A_1 \cup A_2)$. Moreover, A_3 itself has measure 0 or ∞.

In particular, if $f \in \Delta$ and $V(x) = \{\theta\in\Theta | f(\theta) = x\}$, then $\mu(V(x) \cap A_i) = \mu(A_i)$ for a unique value of $x \geq 0$, say x_i, unless

$\mu(A_i) = 0$. Consequently, any f in Δ can be written as $x_1 \cdot I(A_1) + x_2 \cdot I(A_2)$, i.e. (Θ,μ) is equivalent to (Θ_2,ν). ∎

Evidently, intangible spaces are rather sparse. They are, however, of some interest and practical importance, as the reader can judge from Examples 2.5.5 and 3.1.3, say. If a space (Θ,μ) is intangible and includes two sets A_1,A_2 with properties (I) and (II), we shall say that it is <u>dichotomous</u>.

With these definitions, we are now in a position to give a precise statement of McConway's result (the proof appears below):

<u>Theorem 2.2.4</u> (McConway 1981)
Let (Θ,μ) be a tangible measurable space. A pooling operator $T:\Delta^n \longrightarrow \Delta$ has the SSFP iff

$$T(f_1,\ldots,f_n) = \sum_{i=1}^{n} w_i f_i \qquad \mu\text{-a.e.}$$

for some $w_i \geq 0$, $\sum_{i=1}^{n} w_i = 1$.

McConway observes that the restriction to what we call tangible spaces (Θ,μ) is hardly relevant to his argument because if Θ had only two neighbourhoods, no nontrivial marginalization could be performed. In view of Lemma 2.2.3 above, this is transparent. That Theorem 2.2.4 cannot be generalized to intangible spaces is illustrated by the following

Example 2.2.5

Suppose $\Theta = \{\theta_1, \theta_2\}$ and μ = counting measure.

Let $G:[0,1]^2 \longrightarrow [0,1]$ be such that

$$G(x,y) = \begin{cases} 0 \text{ if } x<y; \\ x \text{ if } x=y; \\ 1 \text{ if } x>y, \end{cases}$$

and consider $T:\Delta^2 \longrightarrow \Delta$ defined by $T(f_1,f_2)(\theta) = G(f_1(\theta),f_2(\theta))$. Then T has the Strong Setwise Function Property and it is easy to check that there are no weights w_1 and w_2 in $[0,1]$ for which $T(f_1,f_2) = w_1 f_1 + w_2 f_2$ for all $f_1,f_2 \in \Delta$.

Proof of Theorem 2.2.4:

One implication is obvious. To prove the other one, let A_1,A_2,A_3 be three neighbourhoods with properties (I) and (II), and consider

$$f_i = \sum_{j=1}^{3} a_i(j) \cdot I(A_j)$$

where $a_i(1) \cdot \mu(A_1) = x_i$, $a_i(2) \cdot \mu(A_2) = y_i$ and $a_i(3) \cdot \mu(A_3) = 1-x_i-y_i \geq 0$ for given x_i,y_i in $[0,1]$ with $x_i+y_i \leq 1$, $1 \leq i \leq n$. Then (2.2.1) implies

$$\int I(A_1) \cdot T(f_1,\ldots,f_n) d\mu = F(\tilde{x}),$$

$$\int I(A_2)\cdot T(f_1,\ldots,f_n)d\mu = F(\tilde{y}),$$

and also

$$\int I(A_1 \cup A_2)\cdot T(f_1,\ldots,f_n)d\mu = F(\tilde{x}+\tilde{y}),$$

where $\tilde{x}=(x_1,\ldots,x_n)$, $\tilde{y}=(y_1,\ldots,y_n)$ and $\tilde{x}+\tilde{y}=(x_1+y_1,\ldots,x_n+y_n)$. Therefore F satisfies Cauchy's functional equation (2.1.1). Using Lemma 2.1.2, it follows that $F(x_1,\ldots,x_n) = \sum_{i=1}^{n} w_i x_i$ on $[0,1]^n$ for some $w_1,\ldots,w_n$ in R. As F must be non-decreasing in each of its components, $w_i \geq 0$ for all $i=1,\ldots,n$; moreover, the fact that $T(f_1,\ldots,f_n)$ is always normalized forces $\sum_{i=1}^{n} w_i = 1$. ■

At first sight, MP seems sensible. It is a principle which guards against inconsistency of probability assessments when performing a "marginal analysis." The necessary pooling can be accomplished either before or after the marginals are reported; the same result is obtained by either route. However, there are two typical situations in which the need for a marginal analysis will arise, and they both cast doubt on the validity of the MP principle:

Case I: Θ is a product parameter space over which each member of a group of experts has been asked to produce his "multivariate" distribution. However, the decision maker is only interested in a particular variable. In that case, McConway explains that MP appears counter-intuitive, at least if the assessors are experts in differing fields and hence have disparate prior knowledge. It could be, for instance, that only one assessor has specialized knowledge in the variable of interest.

Case II: Θ represents only one assessment situation and each member of the panel has suggested his "univariate" distribution to the decision maker. In that case, when would the decision maker feel compelled to reduce the size of $\Omega(\mu)$, the class of all possible "events"? The answer is: in the light of new evidence revealed to him (and to the panel) to the effect that certain alternatives which had been considered possible a priori can now be ruled out because they have become "asserted logical impossibilities" (Koopman 1940). Upon being apprised of this new information, he would update his beliefs using Bayes' Theorem, the only rational prescription for updating probability distributions, subjective or not. [1]

Given that in general the processes of marginalizing and updating priors using Bayes' rule are not equivalent, what would seem to be called for, here, is not so much McConway's MP

[1] In this regard, cf. French (1982) who provides axioms justifying the use of Bayes' Theorem when "changes of information take the form of the occurence of an event in the field upon which the subject is concentrating."

condition as an axiom which would guarantee that the same final consensus distribution is arrived at whether (i) the experts' priors are combined first and the resulting consensus opinion is updated; or else (ii) the posteriors are derived by each expert individually and then pooled by the decision maker. This axiom already exists, and Madansky (1964,1978) calls it External Bayesianity. For a more technical definition of this concept, as well as an analysis of some of its consequences, the reader is referred to Section 2.4 below.

Focusing on (2.2.1) now, a condition which is essentially equivalent to MP, we see that the probability assigned by the consensus distribution, $T(f_1,\ldots,f_n)$, to any measurable set $A \in \Omega(\mu)$ is assumed to depend solely on the probabilities given to A by the individual assessors' distributions $f_1,\ldots,f_n$. As $\Omega(\mu)$ is potentially very large and A is arbitrary in $\Omega(\mu)$, this condition is obviously a far-reaching one. Indeed, not only does (2.2.1) dictate the local behaviour of $T(f_1,\ldots,f_n)$ in terms of the f_i's (A's with $0 < \mu(A) << 1$), it also controls its global behaviour (A's with $\mu(A) >> 1$) at the same time. We think that this is unnecessary and that it should suffice to examine what $T(f_1,\ldots,f_n)$ does on the atoms of μ (for definition, cf. Royden 1968, p. 321). Thus, in the following section, our first task will consist of characterizing the linear opinion pool by assuming only what we have earlier defined as "locality" (cf. Definition 2.1.1), a condition which amounts roughly to (2.2.1) restricted to atoms.

2.3 A characterization of the linear opinion pool via locality

The purpose of this section is to establish that the linear opinion pool is the only local pooling operator which preserves dogmas. Both definitions were given in Section 2.1. The term "locality" mimics Bernardo (1979) who uses it to describe an analogous property of utility functions.

Roughly speaking, locality reduces pooling operators to Lebesgue-measurable functions G on $[0,\infty)^n$, and thus constitutes a fairly strong requirement. But it is intuitively appealing and certainly constitutes a viable alternative to McConway's SSFP. It could be viewed as a likelihood principle for pooling operators in the following sense: a particular value of $\theta \epsilon \Theta$ is correct and the consensus probability at θ is required to depend only upon the probabilities assigned to the "true state," and not upon the probabilities of those states of nature which could have obtained but did not. The condition that G be Lebesgue-measurable is not needed explicitly here, but it ties the present material to that of the ensuing sections.

We say that a pooling operator preserves dogmas if it defines a consensus density which is zero on that part of the space where the probability assessors all said that it should be zero. The word "dogma" is borrowed from Bacharach (1973) who observes that if the probability of a certain event was assessed to be zero by an individual, no pertinent evidence about that event will ever affect his opinion. His judgement, shall we say, rules out its force in advance, i.e. it is dogmatic. We do not wish to debate, at this point, the problem of whether professing dogmas clashes with the putatively scientific attitude of leaving all questions open to be decided by "facts." However, it seems reasonable to expect that a decision maker who has sought the advice of articulate experts would not challenge their common dogmas. Difficulties will arise only if the dogmas expressed are conflicting.

Note that a local pooling operator need not always preserve dogmas; consider for instance the operator $T:\Delta^n \longrightarrow \Delta$ which would map everything to $I(\Theta)/\mu(\Theta)$. However, as will be shown in the next lemma, dogma preservation is automatic when $\mu(\Theta)$ is infinite.

Lemma 2.3.1

Every local pooling operator $T:\Delta^n \longrightarrow \Delta$ preserves dogmas when $\mu(\Theta)$ is infinite.

Proof:

Let $G:[0,\infty)^n \longrightarrow [0,\infty)$ be the Lebesgue-measurable function whose existence is guaranteed by Definition 2.1.1. If $A \epsilon \Omega(\mu)$ is such that $\mu(A) = K$ for some real number $0<K<\infty$, let $f = I(A)/K \epsilon \Delta$ and observe that $1 = \int T(f,\ldots,f)d\mu = KG(1/K,\ldots,1/K) + G(0,\ldots,0)\cdot\mu(\Theta\backslash A) = 1$. Since $G(\tilde{o})\cdot\mu(\Theta\backslash A)$ is finite and $\mu(\Theta\backslash A)$ is infinite, we conclude that $G(\tilde{o}) = 0$, i.e. that T is dogma preserving. ∎

We will use methods from the theory of functional equations to prove:

<u>Theorem 2.3.2</u>

Let (Θ,μ) be tangible. The linear opinion pool is the only local pooling operator which preserves dogmas.

We split the proof of Theorem 2.3.2 into two lemmas.

<u>Lemma 2.3.3</u>

Let A_1,A_2,A_3 be three μ-measurable neighbourhoods in Θ with properties (I) and (II). If $T:\Delta^n \longrightarrow \Delta$ preserves dogmas and there exists a Lebesgue-measurable function $G:[0,\infty)^n \longrightarrow [0,\infty)$ such that

$$T(f_1,\ldots,f_n) = G\Box(f_1,\ldots,f_n) \quad \mu\text{-a.e.},$$

then $G(\tilde{x}) = \sum_{i=1}^{n} w_i x_i$ for all $\tilde{x} \epsilon [0,1/M]^n$, where $M = \min\{\mu(A_1), \mu(A_2),\mu(A_3)\}$.

<u>Proof</u>:

Call $m_i = \mu(A_i)$, $i=1,2,3$ and suppose that $M = m_3$. There is no loss of generality in assuming A_1, A_2, A_3 disjoint: $i \neq j \Rightarrow A_i \cap A_j = \emptyset$. Consider

$$f_i = \sum_{j=1}^{3} a_i(j) \cdot I(A_j)$$

where $a_i(j) \geq 0$ and $\sum_{j=1}^{3} a_i(j) \cdot m_j = 1$ for all $1 \leq i \leq n$. As T preserves dogmas, $G(\tilde{o}) = 0$ and furthermore

$$\int T(f_1, \ldots, f_n) d\mu = \sum_{j=1}^{3} G(\tilde{a}(j)) \cdot m_j = 1 \qquad (2.3.1)$$

where $\tilde{a}(j) = (a_1(j), \ldots, a_n(j))$, $j=1,2,3$.

Define $h:[0,1]^n \longrightarrow [0,1]$ by

$$h(\tilde{c}) = 1 - M \cdot G((1-c_1)/M, \ldots, (1-c_n)/M)$$

for all $\tilde{o} \leq \tilde{c} \leq \tilde{1}$, so that $h(\tilde{o}) = 0$ and $h(\tilde{1}) = 1$.

It will suffice to show that h satisfies Cauchy's functional Equation (2.1.1), for then Lemma 2.1.2 will imply that $h(\tilde{c}) = \sum_{i=1}^{n} w_i c_i$ for some $w_i \geq 0$, and $\sum_{i=1}^{n} w_i = 1$ because $h(\tilde{1}) = 1$.

To establish this fact, note that by Equation (2.3.1),

$$G(\tilde{a}(1)) \cdot m_1 + G(\tilde{a}(2)) \cdot m_2 = h(\tilde{c}) \qquad (2.3.2)$$

whenever $\tilde{a}(3) = (\tilde{1}-\tilde{c})/M$.

In particular, observe that if $\tilde{o} \leq \tilde{x} \leq \tilde{c} \leq \tilde{1}$ are given, the choices $a_i(1) = x_i/m_1$ and $a_i(2) = (c_i - x_i)/m_2$ imply

$$G(\tilde{x}/m_1)\cdot m_1 \quad + \quad G((\tilde{c}-\tilde{x})/m_2)\cdot m_2 \quad = \quad h(\tilde{c}). \qquad (2.3.3)$$

However, taking $a_i(1) = x_i/m_1$ and $a_i(2) = 0$ in (2.3.2) shows that

$$G(\tilde{x}/m_1)\cdot m_1 \quad = \quad h(\tilde{x})$$

and, similarly, choosing $a_i(1) = 0$ with $a_i(2) = (c_i-x_i)/m_2$ establishes that

$$G((\tilde{c}-\tilde{x})/m_2)\cdot m_2 \quad = \quad h(\tilde{c}-\tilde{x}).$$

Consequently, (2.3.3) becomes

$$h(\tilde{c}) \quad = \quad h(\tilde{x}) + h(\tilde{c}-\tilde{x})$$

for all $\tilde{0} \le \tilde{x} \le \tilde{c} \le \tilde{1}$, which we can rewrite as

$$h(\tilde{x}+\tilde{y}) \quad = \quad h(\tilde{x}) + h(\tilde{y})$$

for $\tilde{x}$, $\tilde{y}$ in $[0,1]^n$ with $\tilde{x}+\tilde{y} \in [0,1]^n$ (take $\tilde{y} = \tilde{c}-\tilde{x}$).
This concludes the proof. ∎

Lemma 2.3.4

Suppose that (Θ,μ) is tangible, and let $M_0 = \inf\{\mu(A) | A \in \Omega(\mu) \text{ and } 0<\mu(A)<\infty)\}$. Given $\delta > 0$, it is always possible to find three μ-measurable neighbourhoods A_1,A_2,A_3 which have properties (I) and (II) and are such that $M_0 \le M < M_0 + \delta$, where $M = \min\{\mu(A_1),\mu(A_2),\mu(A_3)\}$.

Proof: Since (Θ,μ) is tangible, there exist at least three μ-measurable neighbourhoods A_1,A_2,A_3 with properties (I) and (II). In fact, we might as well assume that they are disjoint. Let $\delta > 0$ be given; we distinguish two cases:

<u>Case I</u>: $M_0 > 0$

Choose $B \in \Omega(\mu)$ to be such that $M_0 \le \mu(B) < M_0 + \min\{\delta, M_0\}$ and look at $B_i = B \cap A_i$, $i=1,2,3$. Then $\mu(B_i)=0$ or $\ge M_0$ by the definition of M_0, and there can be at most one i, say i=1, such that $\mu(B_i) \ge M_0$, for $2M_0 > \mu(B) \ge \sum_{i=1}^{3} \mu(B_i)$. Our three sets are B_1, A_2 and A_3.

<u>Case II</u>: $M_0 = 0$

Pick $B \in \Omega(\mu)$ so that $0 < \mu(B) < \delta$, and, once again, let $B_i = B \cap A_i$, $i=1,2,3$. Put $m = \max\{\mu(B_1), \mu(B_2), \mu(B_3)\}$ and relabel the sets A_i, B_i so that $\mu(B_1) = m$. If m=0, our choice of sets is B, A_2, A_3; if m>0, then replace B by B_1. ∎

Theorem 2.3.2 is an immediate consequence of the two lemmas above, and Example 2.2.5 shows that the tangibility of the space (Θ, μ) is critical. Note also that, contrary to what one might conjecture at first, not every pooling operator is local on an intangible space. An example to this effect follows.

<u>Example 2.3.5</u>

Let $\Theta = \{\theta_1, \theta_2\}$ and μ = counting measure, as in Example 2.2.5 above. Let $T:\Delta^2 \longrightarrow \Delta$ be defined by $T(f_1,f_2)(\theta_1) = f_1(\theta_1) \cdot f_2(\theta_1)$ and $T(f_1,f_2)(\theta_2) = 1 - f_1(\theta_1) \cdot f_2(\theta_1)$. Then T preserves dogmas but is not local.

It is possible to generalize Theorem 2.3.2 in at least one way. In the proposition stated below and proved in the sequel, the definition of locality is relaxed to allow the function G to depend on θ as well as on the values that the densities $f_1, \ldots, f_n$ take at that point. Thus, we consider all pooling operators

$T: \Delta^n \longrightarrow \Delta$ of the form

$$T(f_1, \ldots, f_n)(\theta) = G(\theta, f_1(\theta), \ldots, f_n(\theta)) \quad \mu\text{-a.e.} \tag{2.3.4}$$

for some measurable function $G: \Theta x [0,\infty)^n \longrightarrow [0,\infty)$. A pooling operator which satisfies (2.3.4) will be called <u>semi</u>-<u>local</u>.

<u>Theorem 2.3.6</u>

Let (Θ,μ) be tangible, and suppose that T is a semi-local pooling operator which preserves dogmas. If the $G(\theta,\cdot)$ of Equation (2.3.4) is <u>continuous</u> as a function on $[0,\infty)^n$ for μ-almost all $\theta \in \Theta$, then T is a linear opinion pool.

Theorem 2.3.6 says that the class of semi-local non-local pooling operators is small. Indeed, when Θ is countable, $|\Theta| \geq 3$ and μ is a counting measure, it is in fact empty. For, take $\tilde{x} \in [0,1]^n$ and let θ, η, λ be three elements in Θ. If $f_i(\theta) = x_i = 1-f_i(\lambda)$ and $g_i(\eta) = x_i = 1-g_i(\lambda)$, the facts that $\int G\square(f_1, \ldots, f_n)d\mu = \int G\square(g_1, \ldots, g_n)d\mu = 1$ and $f_i(\lambda) = g_i(\lambda)$,

$i=1,\ldots,n$ together entail $G(\theta,\tilde{x}) = G(\eta,\tilde{x})$. Our proof of Theorem 2.3.6 is an attempt to generalize this argument to arbitrary spaces.

It is conceivable that the extra (continuity) condition on G could be weakened, but we have not attempted to do so. Note that whatever be (Θ,μ), the requirement that T preserves dogmas is necessary to rule out pooling operators which would map every n-tuple of opinions $(f_1,\ldots,f_n)$ to the same fixed μ-density $g \in \Delta$. These operators are worse than dictatorships, since they correspond to the case where the decision maker's mind is made up in advance and "consultation" is conducted for form's sake only. They would therefore seem to be of little interest.

Here again, the result fails to extend to dichotomous spaces:

<u>Example 2.3.7</u>

Let $\Theta = \{\theta_1,\theta_2\}$, μ = counting measure and consider a function $G:\Theta \times [0,1]^2 \longrightarrow [0,1]$ defined by $G(\theta_1,x,y) = x \cdot I\{y \mid 0 \le y \le 1/2 \text{ or } y=1\} + (1-x)\cdot I\{y \mid 0<y<1/2\}$ and $G(\theta_2,x,y) = x\cdot I\{y \mid y=0 \text{ or } 1/2 \le y \le 1\} + (1-x)\cdot I\{y \mid 0<y<1/2\}$. Then $T:\Delta^2 \longrightarrow \Delta$ defined by $T(f_1,f_2)(\theta) = G(\theta,f_1(\theta),f_2(\theta))$ is semi-local and preserves dogmas. However, it is not local.

The gist of the proof of Theorem 2.3.6 is contained in

<u>Lemma 2.3.8</u>

Let A_1, A_2, A_3 be three μ-measurable neighbourhoods in Θ with properties (I) and (II), and let $T:\Delta^n \longrightarrow \Delta$ be a pooling operator which preserves dogmas. If $G:\Theta \times [0,\infty)^n \longrightarrow [0,\infty)$ is a measurable function such that (2.3.4) holds for all choices of $f_1,\ldots,f_n$ in Δ, then there exist $w_1,\ldots,w_n \in [0,1]$ with $\sum_{i=1}^{n} w_i = 1$ for which

$$G(\cdot,\tilde{x}) = \sum_{i=1}^{n} w_i x_i \quad \mu\text{-a.e. on } A_j \tag{2.3.5}$$

for all $\tilde{x} \in [0,1/\mu(A_j)]^n$ and $j=1,2,3$.

<u>Proof</u>: We divide the proof into three parts.

<u>Step 1</u>: Define $f(\tilde{x}) = \int I(A_1)G(\cdot,\tilde{x})d\mu$ for all $\tilde{x} \in [0,1/\mu(A_1)]^n$; we will show that f satisfies Cauchy's functional Equation (2.1.1).

First note that if $f_i = x_i I(A_1) + y_i I(A_2) + z_i I(A_3)$ is in Δ,

$$\int T(f_1,\ldots,f_n)d\mu = \int I(A_1)G(\cdot,\tilde{x})d\mu + \int I(A_2)G(\cdot,\tilde{y})d\mu + \int I(A_3)G(\cdot,\tilde{z})d\mu = 1 \tag{2.3.6}$$

by the fact that T preserves dogmas. Letting

$g_i = y_i I(A_2) + [1-y_i \mu(A_2)]\cdot I(A_3)/\mu(A_3)$ and

$h_i = y_i [\mu(A_2)/\mu(A_1)]\cdot I(A_1) + [1-y_i \mu(A_2)]\cdot I(A_3)/\mu(A_3)$

both in $\Delta, 1 \le i \le n$, we see that

$$\int T(g_1,\ldots,g_n)d\mu =$$

$$\int I(A_2)G(\cdot,\tilde{y})d\mu + \int I(A_3)G(\cdot,[\tilde{1}-\tilde{y}\mu(A_2)]/\mu(A_3))d\mu = 1$$

and also

$$\int T(h_1,\ldots,h_n)d\mu =$$

$$\int I(A_1)G(\cdot,\tilde{y}\mu(A_2)/\mu(A_1))d\mu + \int I(A_3)G(\cdot,[\tilde{1}-\tilde{y}\mu(A_2)]/\mu(A_3))d\mu = 1.$$

Thus

$$\int I(A_2)G(\cdot,\tilde{y})d\mu =$$

$$\int I(A_1)G(\cdot,\tilde{y}\mu(A_2)/\mu(A_1))d\mu = f(\tilde{y}\mu(A_2)/\mu(A_1)). \qquad (2.3.7)$$

Similarly, we find that

$$\int I(A_3)G(\cdot,\tilde{z})d\mu =$$

$$\int I(A_1)G(\cdot,\tilde{z}\mu(A_3)/\mu(A_1))d\mu = f(\tilde{z}\mu(A_3)/\mu(A_1)) \qquad (2.3.8)$$

and so (2.3.6) now reads

$$f(\tilde{x}) + f(\tilde{y}\mu(A_2)/\mu(A_1)) + f(\tilde{z}\mu(A_3)/\mu(A_1)) = 1$$

whenever $\tilde{x}\mu(A_1) + \tilde{y}\mu(A_2) + \tilde{z}\mu(A_3) = 1$.

Relabelling $\tilde{y} = \tilde{y}\mu(A_2)/\mu(A_1)$, $\tilde{z} = \tilde{z}\mu(A_3)/\mu(A_1)$, we have $f(\tilde{x}) + f(\tilde{y}) + f(\tilde{z}) = 1$ for all $\tilde{x},\tilde{y},\tilde{z}$ in $[0,1/\mu(A_1)]^n$ with $x_i + y_i + z_i = 1/\mu(A_1), 1 \le i \le n$.

So if $\tilde{u}$, $\tilde{v}$ are in $[0,1/\mu(A_1)]^n$ with $\tilde{u} + \tilde{v} \in [0,1/\mu(A_1)]^n$, and if $z_i = (1/\mu(A_1))-u_i-v_i, i=1,\ldots,n$, then $f(\tilde{u}+\tilde{v}) + f(\tilde{o}) + f(\tilde{z}) = 1$

and also $f(\tilde{u}) + f(\tilde{v}) + f(\tilde{z}) = 1$. But again, $f(\tilde{o}) = 0$ because T preserves dogmas, and so $f(\tilde{u}+\tilde{v}) = f(\tilde{u}) + f(\tilde{v})$.

According to Lemma 2.1.2 now, there exist $a_1,\ldots,a_n$ in R such that $f(\tilde{x}) = \sum_{i=1}^{n} a_i x_i$ on $[0,1/\mu(A_1)]^n$, and since $f(\tilde{x}) \geq 0$ always, these constants a_i are non-negative.

Furthermore, $f(\tilde{1}/\mu(A_1)) = 1$, so that $\sum_{i=1}^{n} a_i = \mu(A_1)$. Just put $w_i = a_i/\mu(A_1), 1 \leq i \leq n$.

Step 2: We show that $G(\theta,\tilde{x})$ is μ-almost everywhere constant in θ on $A=A_1$. For that, we use the key fact that for any μ-measurable subset A' of A,

$$\int I(A')G(\cdot,\tilde{x})d\mu = \sum_{i=1}^{n} w_i x_i \mu(A').$$

If $\mu(A') = 0$ or $= \mu(A)$, this is obvious, and if $0 < \mu(A') < \mu(A)$, we can apply the above argument to see that

$$\int I(A')G(\cdot,\tilde{x})d\mu = \sum_{i=1}^{n} w'_i x_i \mu(A')$$

and

$$\int I(A'')G(\cdot,\tilde{x})d\mu = \sum_{i=1}^{n} w''_i x_i \mu(A''),$$

where $A'' = A \backslash A'$ and the primes on the w_i's indicate a possible dependence on the set over which $G(\cdot,\tilde{x})$ is integrated. These primes may, in fact, be dropped; for, if

$$g_i = x_i I(A') + [1-x_i \mu(A')]\cdot I(A_2)/\mu(A_2)$$

and

$$h_i = x_i [\mu(A')/\mu(A")]\cdot I(A") + [1-x_i \mu(A')]\cdot I(A_2)/\mu(A_2)$$

are in $\Delta, 1 \le i \le n$, then

$$\int T(g_1,\ldots,g_n)d\mu = \int I(A')G(\cdot,\tilde{x})d\mu + \int I(A_2)G(\cdot,[1-\tilde{x}\mu(A')]/\mu(A_2))d\mu$$

equals 1 and equals

$$\int T(h_1,\ldots,h_n)d\mu = \int I(A")G(\cdot,\tilde{x}\mu(A')/\mu(A"))d\mu$$
$$+ \int I(A_2)G(\cdot,[\tilde{1}-\tilde{x}\mu(A')]/\mu(A_2))d\mu,$$

from which we conclude

$$\int I(A')G(\cdot,\tilde{x})d\mu = \int I(A")G(\cdot,\tilde{x}\mu(A')/\mu(A"))d\mu$$

and in turn

$$\sum_{i=1}^{n} w'_i x_i \mu(A') = \sum_{i=1}^{n} w"_i x_i \mu(A')$$

for every possible choice of $x_1,\ldots,x_n$ in $[0,1/\mu(A)]$. Thus $w'_i = w"_i$ for all $i=1,\ldots,n$, and moreover $w'_i = w_i$ since

$$\int I(A)G(\cdot,\tilde{x})d\mu = \int I(A')G(\cdot,\tilde{x})d\mu + \int I(A")G(\cdot,\tilde{x})d\mu$$

entails

$$\sum_{i=1}^{n} w_i x_i \mu(A) = \sum_{i=1}^{n} w'_i x_i [\mu(A')+\mu(A")],$$

or $\sum_{i=1}^{n} w_i x_i = \sum_{i=1}^{n} w'_i x_i$ for all $x_1,\ldots,x_n$ in $[0,1/\mu(A)]$.

Finally, fix $\tilde{x} \in [0,1/\mu(A)]^n$ and suppose that for some $\delta > 0$, the set $A' = \{\theta\in A | G(\theta,\tilde{x}) > \sum_{i=1}^{n} w_i x_i + \delta\}$ is non-negligible. Then

$$\int I(A')G(\cdot,\tilde{x})d\mu = \sum_{i=1}^{n} w_i x_i \mu(A') > \sum_{i=1}^{n} w_i x_i \mu(A') + \delta\mu(A'),$$

a contradiction. Hence $G(\cdot,\tilde{x}) \le \sum_{i=1}^{n} w_i x_i$ μ-a.e. on A, and a similar argument shows the reverse inequality.

Step 3: We can repeat steps 1 and 2 for A_2 or A_3 instead of A_1, so that

$$G(\cdot,\tilde{x}) = \sum_{i=1}^{n} w_{ij} x_i \quad \mu\text{-a.e.} \tag{2.3.9}$$

for all $\tilde{x} \in [0,1/\mu(A_j)]^n$ and some given constants w_{ij} in $[0,1]$ satisfying $\sum_{i=1}^{n} w_{ij} = 1$, $j=1,2,3$.

But by (2.3.7),

$$\begin{aligned}\int I(A_2)G(\cdot,\tilde{y})d\mu &= \sum_{i=1}^{n} w_{i2} y_i \mu(A_2)\\ &= \int I(A_1)G(\cdot,\tilde{y}\mu(A_2)/\mu(A_1))d\mu\\ &= \sum_{i=1}^{n} w_{i1} y_i \mu(A_2)\end{aligned}$$

for all $\tilde{y} \in [0,1/\mu(A_2)]^n$, so that $w_{i1} = w_{i2}$, $i=1,\ldots,n$. Similarly, $w_{i1} = w_{i3}$ follows from (2.3.8) and so (2.3.9) entails the stated conclusion. ∎

Let £ denote the collection

$\{A_1 \epsilon \Omega(\mu) | A_1, A_2, A_3$ have properties (I)&(II) for some $A_2, A_3 \epsilon \Omega(\mu)\}$

(note that $\pounds \neq \emptyset \iff (\Theta,\mu)$ is tangible). Lemma 2.3.8 can be strengthened in the following way:

Lemma 2.3.9

Let (Θ,μ) be tangible and let $T:\Delta^n \longrightarrow \Delta$ be a dogma preserving semi-local pooling operator. If $G:\Theta x[0,\infty)^n \longrightarrow [0,\infty)$ denotes the measurable function for which (2.3.4) holds, then there exist $w_1, \ldots, w_n \epsilon [0,1]$ satisfying $\sum_{i=1}^{n} w_i = 1$ and such that

$$G(\cdot,\tilde{x}) = \sum_{i=1}^{n} w_i x_i \quad \mu\text{-a.e. on } A,$$

whatever be $\tilde{x} \epsilon [0,1/\mu(A)]^n$ and $A \epsilon \pounds$.

Proof:

Let A and B in $\pounds$ be so that $G(\cdot,\tilde{x}) = \sum_{i=1}^{n} w_i x_i$ μ-a.e. on A and $G(\cdot,\tilde{y}) = \sum_{i=1}^{n} w'_i y_i$ μ-a.e. on B for some $\{w_i\},\{w'_i\}$ in [0,1] sets of weights, each set adding up to 1, and all $\tilde{x}$ in $[0,1/\mu(A)]^n$, $\tilde{y}$ in $[0,1/\mu(B)]^n$. That the $\{w_i\}$ and $\{w'_i\}$ exist follows from Lemma 2.3.8.

If $\mu(A \cap B) > 0$, then clearly $w_i = w'_i$, i=1,...,n. Otherwise, let

A_1 be a non-negligible subset of $A \backslash B$ and pick $A_2, A_3 \in \Omega(\mu)$ so that A_1, A_2, A_3 have properties (I) and (II). Here, we used the fact that (Θ, μ) is tangible. From Lemma 2.3.8 above, we know that

$$G(\cdot, \tilde{x}) = \sum_{i=1}^{n} w_i x_i \quad \mu\text{-a.e. on } A_j, \tag{2.3.10}$$

where $\tilde{x} \in [0, 1/\mu(A_j)]^n$ and $j=1,2,3$.

If $\mu(B \cap A_j) > 0$ for $j=2$ or 3, we are done.

If not, then $\{B, A_2, A_3\}$ constitutes a set with properties (I) and (II) and we employ Lemma 2.3.8 again to conclude that

$$G(\cdot, \tilde{y}) = \sum_{i=1}^{n} w'_i y_i \quad \mu\text{-a.e. on } A_j, \tag{2.3.11}$$

$\tilde{y}$ being arbitrary in $[0, 1/\mu(A_j)]^n$ and $j=2,3$. Pulling (2.3.10) and (2.3.11) together shows that $w'_i = w_i$, $i=1,\dots,n$. ■

We say that a μ-density $f \in \Delta$ is a <u>simple function</u> iff $f = \sum_{i\geq 1} c_i I(A_i)$ μ-a.e. for some $c_i \geq 0$ and a sequence $\{A_i \in \Omega(\mu) \mid i=1,2\dots\}$ of disjoint sets. With this definition, we can state and prove

<u>Proposition 2.3.10</u>

If $T: \Delta^n \longrightarrow \Delta$ is a dogma preserving semi-local pooling operator on a tangible space (Θ, μ), then there exist $w_1, \dots, w_n \in [0,1]$ such that $\sum_{i=1}^{n} w_i = 1$ and

$$T(f_1,\ldots,f_n) = \sum_{i=1}^{n} w_i f_i \quad \mu\text{-a.e.}$$

for all $f_1,\ldots,f_n$ <u>simple functions</u> in Δ.

<u>Proof</u>:

If $f_1,\ldots,f_n \in \Delta$ are simple functions, it is possible to find a sequence $S = \{A_i \in \Omega(\mu) | i=1,2,\ldots\}$ of disjoint sets A_j together with constants $0 \le c_{ij} < \infty$ for which $f_i = \sum_{j\ge 1} c_{ij} I(A_j)$ μ-a.e., $i=1,\ldots,n$. Since (Θ,μ) is tangible, we can assume that $|S| > 2$, so that the A_j's belong to $\pounds$. Thus, by Lemma 2.3.9, there exist $w_1,\ldots,w_n \in [0,1]$ summing up to 1 for which (2.3.5) holds true. Since T is semi-local, $T(f_1,\ldots,f_n) = G(\cdot,c_{1j},\ldots,c_{nj})$ μ-a.e. on A_j for each $j \ge 1$, and observe that $c_{ij}\cdot\mu(A_j) \le \int f_i \, d\mu \le 1$, so that $G(\cdot,c_{1j},\ldots,c_{nj}) = \sum_{i=1}^{n} w_i c_{ij}$ μ-a.e. on A_j by Lemma 2.3.9.

■

The formulation of the following corollary was suggested by Dr. Harry Joe (personal communication).

<u>Corollary 2.3.11</u>

If (Θ,μ) is tangible and μ is both σ-finite and atomic, the linear opinion pool is the only semi-local pooling operator which preserves dogmas.

Proof: If μ is σ-finite and atomic, the collection C of its atoms is at most countable; furthermore, $|C| \geq 3$ from the fact that (Θ,μ) is tangible. So, if we write $C = \{A_i | i=1,2,3...\}$, every function $f \in \Delta$ can be expressed μ-almost everywhere as an infinite sum $\sum_{i\geq 1} c_i I(A_i)$, i.e. Δ consists of simple functions only. Apply Proposition 2.3.10. ∎

In general, it does not necessarily follow from Proposition 2.3.10 that any dogma preserving semi-local pooling operator is local. What is clear, however, is that if T is continuous with respect to the pointwise convergence topology, then Theorem 2.3.6 and Proposition 2.3.10 are equivalent. This regularity condition is secured by requiring G itself to be continuous.

Lemma 2.3.12

Let $T:\Delta^n \longrightarrow \Delta$ be semi-local and let $G:\Theta x[0,\infty)^n \longrightarrow [0,\infty)$ be the corresponding function for which (2.3.4) holds. If $G(\theta,\cdot)$ is continuous as a function on $[0,\infty)^n$ for μ-almost all $\theta \in \Theta$, then

$$\lim_{k\longrightarrow\infty} T(f_{1k},...,f_{nk}) = T(f_1,...,f_n)$$

whenever $f_{ik} \longrightarrow f_i$ pointwise μ-a.e. as $k\rightarrow\infty$, $i=1,...,n$.

Proof: Let

$$A = \bigcup_{i=1}^{n} \{\theta\epsilon\Theta | \lim_{k\to\infty} f_{ik}(\theta) \neq f_i(\theta)\},$$

$$B = \bigcup_{k\geq 1} \{\theta\epsilon\Theta | T(f_{1k},\dots,f_{nk})(\theta) \neq G(\theta, f_{1k}(\theta),\dots,f_{nk}(\theta))\},$$

and $C = \{\theta\epsilon\Theta | G(\theta,\cdot)$ is not continuous as a function on $[0,\infty)^n\}$. Let also $D = \{\theta\epsilon\Theta | T(f_1,\dots,f_n)(\theta) \neq G(\theta, f_1(\theta),\dots,f_n(\theta))\}$ and denote by E the μ-negligible set AUBUCUD. For all $\theta \;\epsilon\; \Theta\backslash E$ we have

$$\begin{aligned}\lim_{k\to\infty} T(f_{1k},\dots,f_{nk})(\theta) &= \lim_{k\to\infty} G(\theta, f_{1k}(\theta),\dots,f_{nk}(\theta)) \\ &= G(\theta, f_1(\theta),\dots,f_n(\theta)) \\ &= T(f_1,\dots,f_n)(\theta),\end{aligned}$$

i.e. $\lim_{k\to\infty} T(f_{1k},\dots,f_{nk}) = T(f_1,\dots,f_n)$ in the pointwise convergence topology. ∎

To complete the proof of Theorem 2.3.6, it suffices to combine Proposition 2.3.10 with the above lemma, keeping in mind that every non-negative measurable function on a space Θ is the limit of some sequence of simple functions (Royden 1968, p. 224).

In conclusion, we have argued that when pooling opinions on Θ, a dominating measure μ will usually impose itself as a natural choice for both the experts and the decision maker. In that case, opinions take the form of densities with respect to μ, and the condition which we called "locality" (or perhaps semi-locality) seems more readily interpretable than McConway's axiom (2.2.1). McConway's characterization of the linear opinion pool can then be reformulated in terms of locality and appears as Theorem 2.3.2. Theorem 2.3.6 extends this result to so-called "semi-local" pooling operators. These findings will have an important consequence in the following section, where Madansky's idea of External Bayesianity will be studied at some length.

2.4 Seeking Externally Bayesian procedures

In Section 2.2, we suggested that External Bayesianity seemed a more appropriate criterion for selecting pooling formulas than McConway's Marginalization Postulate. In the present section, we will give a precise definition of this concept and investigate some of its implications. In particular, conditions will be stated under which External Bayesianity characterizes the logarithmic opinion pool (2.4.2).

External Bayesianity (EB) has been introduced by Madansky (1964;1978) as an axiom of group rationality for solving decision-making problems. The concept, however, can be readily interpreted within our framework for the problem of the panel of experts. Basically, if the panel were to use an Externally Bayesian procedure to determine a consensus, they would be perceived as acting in the manner of a single Bayesian. This entails updating their beliefs in accordance with Bayes' rule. To insure that they would act in a consistent fashion, it is necessary that the pooling procedure yield the same result whether they pool before or after updating their beliefs in the light of new information.

More precisely, we have the following

Definition 2.4.1

Let $T:\Delta^n \longrightarrow \Delta$ be a pooling operator. We say that T is Externally Bayesian iff

$0 < \int \Phi T(f_1,\ldots,f_n)d\mu < \infty$, and

$$T[\Phi f_1/\int\Phi f_1 d\mu,\ldots,\Phi f_n/\int\Phi f_n d\mu] =$$

$$\Phi T(f_1,\ldots,f_n)/\int\Phi T(f_1,\ldots,f_n)d\mu \quad \mu\text{-a.e.} \qquad (2.4.1)$$

whenever $\Phi:\Theta \longrightarrow [0,\infty)$ is a μ-measurable function such that $0 < \int\Phi f_i d\mu < \infty$ for each $1\le i\le n$ (such a function Φ is called a likelihood function).

Examples of Externally Bayesian procedures are dictatorships and (provided it is well defined) the logarithmic opinion pool,

$$T(f_1,\ldots,f_n) = \prod_{i=1}^{n} f_i^{w(i)} / \int \prod_{i=1}^{n} f_i^{w(i)} d\mu, \quad \sum_{i=1}^{n} w(i)=1. \qquad (2.4.2)$$

In his book on decision analysis, Raiffa (1968) illustrates what can happen if the processes of updating and pooling probability distributions do not commute. He gives an example (on a dichotomous space) in which two experts find it in their own best interest to convince the decision maker to compute the consensus distribution <u>before</u> he learns of the outcome of an experiment. They do so in order to maximize the impact of their opinions on the consensus perceived by the decision maker, regardless of the outcome of the experiment. Such behaviour need not be entirely selfish and motivated only by the desire "to win." In case they disagree, it would be quite reasonable to expect that each expert would believe he is right. However, new, relevant evidence should always be welcomed -by both the experts and the decision maker- and the question of whether to update opinions before or after a consensus is found should not admit the possibility of the experts gaining some advantage for their opinions over the new evidence by strategic manoeuvring.

External Bayesianity has also been advocated by Weerahandi & Zidek (1978) who call it "prior-to-posterior coherence." Their rationale for using this axiom derives from the observation that if each expert is a Bayesian, his prior opinion might well have been his posterior in an earlier experiment, and that similarly, he will use the posterior which will result from his present investigations as his future prior.

Thus, all in all, External Bayesianity seems to be an eminently reasonable prescription for selecting "good" pooling operators. We commence our analysis of its implications with an easy lemma.

Lemma 2.4.2

Let $T:\Delta^n \longrightarrow \Delta$ be an Externally Bayesian pooling operator. Then T preserves dogmas and furthermore $T(f_1,\ldots,f_n) = T(g_1,\ldots,g_n)$ μ-a.e. whenever $f_i = g_i$ μ-a.e. for all $1 \le i \le n$.

Proof:

Let $f_1,\ldots,f_n \in \Delta$ be such that $Z = \{\theta\epsilon\Theta | f_1(\theta)=\ldots=f_n(\theta)=0\}$ is non-negligible (i.e. $\mu(Z)>0$). If $\Phi = I(\Theta\backslash Z)$, then $\Phi f_i = f_i$ and so $\int\Phi f_i \, d\mu = 1$, $1\le i\le n$. Using Equation (2.4.1), it follows that $\int\Phi T(f_1,\ldots,f_n)d\mu = K$ for some real number $0<K<\infty$ and also

$$T(f_1,\ldots,f_n) = \Phi T(f_1,\ldots,f_n)/K \quad \mu\text{-a.e.}$$

But the right-hand side equals 0 μ-a.e. on Z, so that

$\mathrm{Supp}(T(f_1,\ldots,f_n)) \subset \bigcup_{i=1}^{n} \mathrm{Supp}(f_i)$, i.e. T is dogma preserving.

To prove the second assertion, suppose that $f_i = g_i$ μ-a.e. and let $A_i = \{\theta\epsilon\Theta | f_i(\theta) = g_i(\theta)\}$ in $\Omega(\mu)$, i=1,2,...,n. Define $A = \bigcup_{i=1}^{n} A_i$ and $\Phi = I(A)$.

Since $\mu(A) = 0$, $\Phi f_i = f_i$ μ-a.e. and similarly $\Phi g_i = g_i$ μ-a.e. Consequently, $\int \Phi f_i \, d\mu = \int \Phi g_i \, d\mu = 1$, and furthermore

$$\int \Phi T(f_1, \ldots, f_n) d\mu = \int \Phi T(g_1, \ldots, g_n) d\mu = 1.$$

And now, using the hypothesis that T is Externally Bayesian, we find that

$$T(\Phi f_1, \ldots, \Phi f_n) = \Phi \cdot T(f_1, \ldots, f_n) \quad \mu\text{-a.e.}$$

and also

$$T(\Phi g_1, \ldots, \Phi g_n) = \Phi \cdot T(g_1, \ldots, g_n) \quad \mu\text{-a.e.}$$

However, $\Phi f_i = \Phi g_i$ everywhere, and hence

$$T(f_1, \ldots, f_n) = T(g_1, \ldots, g_n) \quad \mu\text{-a.e.} \blacksquare$$

Section 2.2 above conveyed our view that new information will more realistically cause a panel of experts to update their probability distributions via Bayes' Theorem than to marginalize them, although both procedures will sometimes yield the same answer. Had McConway postulated External Bayesianity instead of his Marginalization Postulate, he would have obtained a very different result:

<u>Theorem 2.4.3</u> (An Impossibility Theorem)
Let (Θ, μ) be tangible. The only Externally Bayesian local pooling operators are dictatorships.

Proof:

Let $T:\Delta^n \longrightarrow \Delta$ be local and Externally Bayesian. Then we know from Lemma 2.4.2 that T also preserves dogmas, and hence

$T(f_1,\ldots,f_n) = \sum_{i=1}^{n} w_i \cdot f_i$ μ-a.e. for some $w_i \geq 0$, $\sum_{i=1}^{n} w_i = 1$ as a consequence of Theorem 2.3.2. We show that $w_j = 1$ for some $j=1,\ldots,n$.

Let $A_1, A_2 \in \Omega(\mu)$ have properties (I) and (II). Such sets exist because (Θ,μ) is tangible, and we can take them to be disjoint. Now pick $i \neq j$ in $\{1,\ldots,n\}$ and consider $f_i = I(A_1)/\mu(A_1)$, $f_k = I(A_2)/\mu(A_2)$ where k runs over the set of indices $\{1,\ldots,n\}\setminus\{i\}$. If $\Phi = x \cdot I(A_1) + y \cdot I(A_2)$ for some x and y in $(0,\infty)$, $x \neq y$, then Equation (2.4.1) applied on A_1 implies that

$$w_i/x = w_i/[w_i \cdot x + (1-w_i)\cdot y].$$

Assuming that w_i is neither 0 nor 1, we conclude that $x=y$, a contradiction. ■

Remark 2.4.4

Under the hypotheses of Theorem 2.3.6, the above result also holds for semi-local pooling operators. Note that the condition that (Θ,μ) be tangible is indispensable, as evidenced by the pooling operator of Example 2.2.5.

This theorem generalizes Genest (1982) and conflicts with a previous finding of Weerahandi & Zidek (1978). In their manuscript, these authors proposed a derivation of the logarithmic opinion pool (2.4.2) based both on External Bayesianity and locality. In view of Theorem 2.4.3, this is only true when all w_i's are 0 but one, and we interpret any function raised to the power 0 as the characteristic function $I(\Theta)$ of the whole space.

We call Theorem 2.4.3 an "Impossibility Theorem" to emphasize that dictatorships of opinions cannot generally be regarded as desirable. Indeed, we would be inclined to follow Bacharach's (1975) policy on this matter and make dictatorships inadmissible. In this case, the theorem would read: "there are no Externally Bayesian local pooling operators."

Next, we extend our search for Externally Bayesian pooling operators to the class of <u>quasi-local</u> procedures, i.e. operators of the form

$$T(f_1,\ldots,f_n) = G\square(f_1,\ldots,f_n)/\int G\square(f_1,\ldots,f_n)d\mu \quad \mu\text{-a.e.} \qquad (2.4.3)$$

where $G:[0,\infty)^n \longrightarrow [0,\infty)$ is a Lebesgue-measurable function with the rather distinctive property that

$$0 < \int G\square(f_1,\ldots,f_n)d\mu < \infty \qquad (2.4.4)$$

for all choices of $f_1,\ldots,f_n$ in Δ. This definition of quasi-locality is equivalent to that given at the beginning of this chapter (Definition 2.1.1). Note that G is not unique, as we could multiply top and bottom of the right-hand side of (2.4.3) by any non-zero positive constant without altering T.

We have already encountered one Externally Bayesian quasi-local pooling operator, namely the logarithmic pooling formula (2.4.2). Here, $G(\tilde{x}) = \prod_{i=1}^{n} x_i^{w(i)}$ and if $\Phi:\Theta \longrightarrow [0,\infty)$ is such that $0 < K_i = \int\Phi f_i d\mu < \infty$, then

$$\begin{aligned} T(\Phi f_1/K_1,\ldots,\Phi f_n/K_n) &= \prod_{i=1}^{n} [\Phi f_i/K_i]^{w(i)} / \int \prod_{i=1}^{n} [\Phi f_i/K_i]^{w(i)} d\mu \\ &= \Phi \prod_{i=1}^{n} f_i^{w(i)} / \int\Phi \prod_{i=1}^{n} f_i^{w(i)} d\mu \\ &= \Phi T(f_1,\ldots,f_n)/\int\Phi T(f_1,\ldots,f_n)d\mu \end{aligned}$$

provided $\sum_{i=1}^{n} w(i) = 1$. In order to ensure that Condition (2.4.4) is met, however, it is necessary to <u>restrict the domain of T</u> to a smaller class of μ-densities $f_1,\ldots,f_n$ for which the integral $\int \prod_{i=1}^{n} f_i^{w(i)} d\mu$ is strictly positive (that $\int \prod_{i=1}^{n} f_i^{w(i)} d\mu$ is always finite follows from Hölder's inequality, at least when the w(i)'s are non-negative; cf. Marshall & Olkin 1979, p. 457).

Here, we have chosen to use

$$\Delta_0 = \{f\epsilon\Delta | f\neq 0 \ \mu\text{-a.e.}\}$$

both for simplicity and ease of exposition. If $\Delta_0 \neq \emptyset$, our analysis suggest a fair amount about the behaviour of Externally Bayesian quasi-local pooling operators acting on Δ^n. In fact, knowing from Lemma 2.4.2 that quasi-local Externally Bayesian procedures preserve dogmas, the only situation which the restriction to Δ_0 fails to encompass is that where some event E in $\Omega(\mu)$ would have been deemed "impossible" (zero probability) by some of the experts but not by all. That this occasion should arise after the experts exchanged their views (as we have assumed they have) is unlikely. Moreover, it is unrealistic to expect the decision maker to reconcile the irreconcilable. This somewhat pathological situation is indeed not unlike that faced in conventional Bayesian analysis when the prior and likelihood functions have disjoint support and some improvisation is called for.

The problem which we will now address is: are there any Externally Bayesian quasi-local pooling operators $T:\Delta_0^n \longrightarrow \Delta_0$ besides (2.4.2)? The answer is <u>no</u>, at least when one is willing to make an extra assumption about (Θ,μ), namely

Assumption 2.4.5

There exist non-negligible μ-measurable sets in Θ of arbitrary small measure, i.e.

$$\forall \delta \in (0,\infty)\ \exists A \in \Omega(\mu) \text{ such that } 0 < \mu(A) < \delta. \qquad (2.4.5)$$

Indeed, we will now prove the following characterization of the logarithmic pooling operator:

Theorem 2.4.6

Suppose (Θ,μ) satisfies Assumption 2.4.5. The logarithmic opinion pool (2.4.2) is the only Externally Bayesian quasi-local pooling operator $T:\Delta_0^n \longrightarrow \Delta_0$.

Remark 2.4.7

If (Θ,μ) satisfies Assumption 2.4.5, then clearly it is tangible and Θ is infinite. Thus Theorem 2.4.6 above does not cover the important case where Θ is finite. The answer in the latter case is unknown.

A special case of the following lemma will prove useful in establishing Theorem 2.4.6:

Lemma 2.4.8

Suppose (Θ,μ) satisfies Assumption 2.4.5. Given $\delta > 0$, there exists a sequence $\{A_n \in \Omega(\mu) | n=1,2,\ldots\}$ of mutually disjoint sets such that $0 < \mu(A_n) < \delta$ for all $n \geq 1$.

Proof:

The proof is by induction. If $\delta > 0$ is given and $A_1,\ldots,A_n$ are n mutually disjoint μ-measurable neighbourhoods with $0 < \mu(A_i) < \delta$, $i=1,\ldots,n$, let $B \in \Omega(\mu)$ such that

$$0 < \mu(B) < (1/2)\cdot\min\{\mu(A_i)\,|\,1\leq i\leq n\} < \delta.$$

Then $\mu(B) < \delta$, $\mu(A_i\backslash B) \geq \mu(A_i)/2$ and so $\{A_1\backslash B,\ldots,A_n\backslash B,B\}$ forms a collection of $n+1$ mutually disjoint sets in $\Omega(\mu)$. ∎

Another obvious consequence of Assumption 2.4.5 is that the function G in (2.4.3) must be defined everywhere on $(0,\infty)^n$:

Lemma 2.4.9

Suppose (Θ,μ) satisfies Assumption 2.4.5, and let $x_1,\ldots,x_n$ be given in $(0,\infty)$. There exist $f_1,\ldots,f_n$ in Δ_0 such that $\mu(\bigcap_{i=1}^{n}\{\theta\in\Theta\,|\,f_i(\theta)=x_i\}) > 0$.

Proof:

Write $\delta = \min\{1/x_i\,|\,1\leq i\leq n\}$ and use Lemma 2.4.8 to choose $A \in \Omega(\mu)$ such that $0 < \mu(A) < \delta$ and $\mu(\Theta\backslash A) > 0$. If $h \in \Delta_0$ is any given μ-density, then $\int hI(\Theta\backslash A)d\mu > 0$ for otherwise h would vanish on some set of strictly positive measure, a contradiction.

Define

$$f_i = x_i \cdot I(A) + p_i \cdot hI(\Theta\backslash A)$$

where $p_i = [1-x_i \mu(A)]/\int hI(\Theta\backslash A)d\mu$, $i=1,\ldots,n$. Clearly $f_i \in \Delta_0$

and $\bigcap_{i=1}^{n} \{\theta\in\Theta | f_i(\theta)=x_i\} = A$ is non-negligible. ∎

We start the proof of Theorem 2.4.6 with

<u>Proposition 2.4.10</u>

Suppose (Θ,μ) satisfies Assumption 2.4.5. If $T:\Delta_0^n \longrightarrow \Delta_0$ is Externally Bayesian and of the form (2.4.3) for some Lebesgue-measurable $G:(0,\infty)^n \longrightarrow (0,\infty)$, then $G(c\tilde{x}) = c\cdot G(\tilde{x})$ for all $c\geq 0$ and $\tilde{x} \in (0,\infty)^n$.

<u>Proof</u>:

If $c=0$, then $G(\tilde{o})=0$ by Lemma 2.4.2 (T preserves dogmas). So suppose $c>0$ and let $\tilde{x}>\tilde{o}$ be fixed.

Given $\delta = \min\{[x_i(c+1)]^{-1} | 1\leq i\leq n\}$, we can use Lemma 2.4.8 to find five disjoint elements A,B,C,D,E of $\Omega(\mu)$ with measure in $(0,\delta)$. Let $\gamma > 0$ be such that

$$\gamma < \min\{2[x_i^{-1} - \mu(A) - c\mu(B)]/\mu(A\cup B)\}$$

and pick $0 < \lambda, \xi < \infty$ so that

$$\lambda < \min\{-\gamma+2\cdot[x_i^{-1}-\mu(A)-c\mu(B)]/\mu(A\cup B) \mid 1\le i\le n\}$$

$$\le \max\{-\gamma+2\cdot[x_i^{-1}-\mu(A)-c\mu(B)]/\mu(A\cup B) \mid 1\le i\le n\} < \xi.$$

Now for each $1\le i\le n$, there exists $d_i \in (0,1)$ so that

$$\lambda d_i + \xi(1-d_i) = -\gamma + 2\cdot[x_i^{-1}-\mu(A)-c\mu(B)]/\mu(A\cup B).$$

Define f_i as

$$I(A\cup B)/2\mu(A\cup B) + d_i\cdot I(C)/4\mu(C) + (1-d_i)\cdot I(D)/4\mu(D) + h\cdot I(N)/4S,$$

where $N = \Theta\backslash(A\cup B\cup C\cup D)$, $S = \int hI(N)d\mu$, and h is some arbitrary function in Δ_0. Note that, here again, $\int h\cdot I(N)d\mu > 0$ for otherwise h would vanish on E, a set of strictly positive measure. It is easy to check that f_i belongs to Δ_0.

Now consider

$$\Phi` = I(A) + cI(B) + \lambda I(C) + \xi I(D) + \gamma I(N).$$

We have that $\Phi\cdot f_i \ne 0$ μ-a.e. and

$$\int\Phi\cdot f_i d\mu = [\mu(A)+c\mu(B)]/2\mu(A\cup B) + \lambda d_i/4 + \xi(1-d_i)/4 + \gamma/4$$

$$= [2x_i\mu(A\cup B)]^{-1} = K_i.$$

Write $u = 1/2\mu(A\cup B)$, so that $u = x_iK_i$, $1\le i\le n$. Now T is Externally Bayesian, i.e.

$$\frac{G(\Phi f_1/K_1,\ldots,\Phi f_n/K_n)}{\Phi G(f_1,\ldots,f_n)} = \frac{\int G(\Phi f_1/K_1,\ldots,\Phi f_n/K_n)\,d\mu}{\int \Phi G(f_1,\ldots,f_n)\,d\mu} \tag{2.4.6}$$

Observe that the right-hand side of this expression is a constant independent of the set (A,B,C,D or N) on which both Φ and the f_i's are evaluated. So, in particular, the left-hand side is the same whether on A or on B. Hence

$$G(u/K_1,\ldots,u/K_n) = (1/c)\cdot G(cu/K_1,\ldots,cu/K_n)$$

upon cancelling a common factor of $G(u,\ldots,u)$ on both sides of the equation. Recalling the definition of u and of the K_i's, we find that

$$c\cdot G(x_1,\ldots,x_n) = G(cx_1,\ldots,cx_n),$$

as asserted in the statement of the proposition. ∎

Thus if a pooling operator $T:\Delta_0^n \longrightarrow \Delta_0$ is both quasi-local and Externally Bayesian, Proposition 2.4.10 above tells us that its corresponding G must be at least "homogeneous." However, the technique which we have used to reach this conclusion could not be applied successfully in cases when G need not be defined over the entirety of $(0,\infty)^n$, as when Θ is finite for example.

Not all homogeneous G's generate an Externally Bayesian quasi-local pooling operator. Consider for instance the function $G(\tilde{x}) = \max\{x_i \mid 1\le i\le n\}$, which gives rise to the quasi-local procedure

$$T(f_1,\ldots,f_n) = \max\{f_1,\ldots,f_n\}/\int\max\{f_1,\ldots,f_n\}d\mu.$$

Clearly G is homogeneous, but T is not Externally Bayesian, as Proposition 2.4.11 will now establish.

Proposition 2.4.11

Let $\tilde{1}$ be a vector of ones and write $\tilde{x}\cdot\tilde{y}$ for the vector $(x_1y_1,\ldots,x_ny_n)$. Then $G(\tilde{x})\cdot G(\tilde{y}) = G(\tilde{x}\cdot\tilde{y})\cdot G(\tilde{1})$ for all $\tilde{x},\tilde{y}$ vectors in $(0,\infty)^n$, where G is the function specified in Equation (2.4.3).

Proof: Let

$$0 < \gamma < \min\{1, x_i^{-1} \mid 1\le i\le n\},$$

$$0 < \delta < \min\{(1-\gamma)/y_i, (x_i^{-1}-\gamma)/y_i\},$$

and let A,B,C,D be disjoint elements of $\Omega(\mu)$ with measure in $(0,\delta)$. If we write $t_i = 1-(\gamma+y_i\mu(B))$, then $t_i > 0$ and $1/x_i > \gamma + y_i\mu(B)$ for all $i=1,\ldots,n$.

Next, choose $0 < \lambda,\xi < \infty$ so that

$$\lambda < \min\{t_i^{-1}\cdot[x_i^{-1}-\gamma-y_i\mu(B)] \mid 1\le i\le n\}$$

$$\le \max\{t_i^{-1}\cdot[x_i^{-1}-\gamma-y_i\mu(B)] \mid 1\le i\le n\} < \xi.$$

Then for each $1 \le i \le n$ there exists a unique $d_i \in (0,1)$ such that

$$\lambda d_i + \xi(1-d_i) = t_i^{-1}[x_i^{-1} - \gamma - y_i\mu(B)].$$

Define

$$f_i = \gamma I(A)/2\mu(A) + y_i I(B) + t_i d_i I(C)/\mu(C)$$
$$+ t_i(1-d_i)I(D)/\mu(D) + (\gamma/2)\cdot[hI(N)/\int hI(N)d\mu]$$

where $h \in \Delta_0$ is arbitrary, and $N = \Theta\backslash(A \cup B \cup C \cup D)$. (That $\int hI(N)d\mu$ is not 0 is a consequence of Lemma 2.4.8.)

Now $f_i \neq 0$ μ-a.e. and

$$\int f_i d\mu = \gamma + y_i\mu + t_i = 1,$$

and hence $f_i \in \Delta_0$, $1 \le i \le n$.

Consider $\Phi = I(A \cup B) + I(N) + \lambda I(C) + \xi I(D)$; we have

$$\int \Phi f_i d\mu = \gamma/2 + y_i\mu(B) + t_i d_i \lambda + t_i(1-d_i)\xi + \gamma/2$$
$$= 1/x_i \quad \text{for } 1 \le i \le n,$$

and since $\Phi f_i \neq 0$ μ-a.e., we may use the fact that T is Externally Bayesian to deduce that the left-hand side of Equation (2.4.6) remains constant as the f_i's and Φ are evaluated on A and B respectively. Consequently, we find

$$\frac{G(\beta x_1,\ldots,\beta x_n)}{G(\beta,\ldots,\beta)} = \frac{G(x_1y_1,\ldots,x_ny_n)}{G(y_1,\ldots,y_n)}$$

where $\beta = \gamma/2\mu(A)$. But by Proposition 2.4.10, the left-hand side reduces to $G(x_1,\ldots,x_n)/G(1,1,\ldots,1)$, whence the result. ■

Proof of Theorem 2.4.6:

Consider $H(\tilde{x}) = G(\tilde{x})/G(\tilde{1})$, a function defined on $(0,\infty)^n$. Then H is Lebesgue-measurable and it follows from Proposition 2.4.11 that $H(\tilde{x}\cdot\tilde{y}) = H(\tilde{x})\cdot H(\tilde{y})$ on its domain. By Lemma 2.1.3, we conclude to the existence of n real numbers $w(1),\ldots,w(n)$ such that $H(\tilde{x}) = \prod_{i=1}^{n} x_i^{w(i)}$ always, i.e.

$$G(\tilde{x}) = G(\tilde{1})\cdot \prod_{i=1}^{n} x_i^{w(i)}.$$

Therefore

$$T(f_1,\ldots,f_n) = \prod_{i=1}^{n} f_i^{w(i)} \Big/ \int \prod_{i=1}^{n} f_i^{w(i)}\, d\mu \qquad \mu\text{-a.e.}$$

The fact that $\sum_{i=1}^{n} w(i) = 1$ follows directly from Proposition 2.4.10: if $\tilde{x}>0$ and $c>0$ are given, we have $G(c\tilde{x})/G(\tilde{1}) = c^{\Sigma w(i)} \cdot \prod_{i=1}^{n} x_i^{w(i)}$

$= c^{\Sigma w(i)} \cdot G(\tilde{x})/G(\tilde{1}) = cG(\tilde{x})/G(\tilde{1})$.

This completes the proof of Theorem 2.4.6. ■

Summarizing our investigations on quasi-locality and External Bayesianity, we have seen that <u>provided the class of "admissible" μ-densities is suitably restricted</u>:

(i) the logarithmic opinion pool $\prod_{i=1}^{n} f_i^{w(i)} / \int \prod_{i=1}^{n} f_i^{w(i)} d\mu$ with $\sum_{i=1}^{n} w(i) = 1$ is Externally Bayesian whatever (Θ,μ);

(ii) if (Θ,μ) satisfies Assumption 2.4.5, the logarithmic opinion pool is the <u>only</u> Externally Bayesian quasi-local pooling procedure available.

In fact, this second conclusion can be somewhat strengthened, as we will presently show:

<u>Proposition 2.4.12</u>

If (Θ,μ) is such that $\Omega(\mu)$ contains an infinite sequence $\{A_n | n\geq 1\}$ of mutually disjoint sets, then the logarithmic opinion pooling operator (2.4.2) is not quasi-local unless $w_1,\ldots,w_n$ are taken to be non-negative.

<u>Proof</u>:

It suffices to show that given $a > 0$, we can find $f,g \in \Delta_0$ with $\int f\cdot(f/g)^a d\mu = \infty$. For, if $w_i<0$ for some $i\in\{1,2,\ldots,n\}$, let $a = -w_i>0$ and consider $f_i = g$, $f_j = f$, $j\neq i$, so that $\int f\cdot(f/g)^a d\mu = \infty$, a contradiction.

Use Lemma 2.4.8 to find a sequence $\{A_n|n\geq 1\}$ of disjoint μ-

measurable neighbourhoods, and define

$$f = \sum_{i\geq 1} K_1 I(A_i)/[\mu(A_i)i^2] + K_2 hI(N)/\int hI(N)d\mu$$

for some $h \in \Delta_0$. (If $\int hI(N)d\mu = 0$, then $N = \Theta\backslash(\bigcup_{i\geq 1} A_i)$ has measure zero, in which case let $f = K_1 \sum_{i\geq 1} I(A_i)/[\mu(A_i)i^2]$ instead.)

In order that f be in Δ_0, it is necessary to have $K_1\pi^2/6 + K_2 = 1$, so $K_1, K_2 > 0$ can be chosen accordingly.

Put $g = L_1 \cdot \sum_{i\geq 1} I(A_i)/\mu(A_i)i^c + L_2 hI(N)/\int hI(N)d\mu$ where c equals $2(a+1)/a > 2$. Then $g \geq 0$ μ-a.e. and $\int g d\mu = L_1 \sum_{i\geq 1} 1/i^c + L_2$ can be made equal to 1 with appropriate choices of L_1, L_2, since $0 < \sum_{i\geq 1} 1/i^c < \infty$.

Now $\int f\cdot(f/g)^a d\mu \geq \sum_{i\geq 1} [K_1/\mu(A_i)i^2]^{a+1} \cdot [\mu(A_i)i^c/L_1]^a \mu(A_i) = K_1^{a+1} \cdot L_1^a \cdot \sum_{i\geq 1} 1 = \infty$. ■

In this last proposition, the hypothesis that there be at least countably many disjoint neighbourhoods in Θ is clearly necessary. If (Θ,μ) is finite, there is no reason why some of the w_i's could not be strictly negative, as long as $\sum_{i=1}^{n} w_i = 1$.

We complete this section with an example to show that an

Externally Bayesian operator need not always preserve unanimity.

Example 2.4.13

Let (Θ,μ) be dichotomous or tangible, so that there exist $A_1,A_2 \in \Omega(\mu)$ with properties (I) and (II). Write $A = A_1$ and $B = \Theta\backslash A$, so that $\min\{\mu(A),\mu(B)\} > 0$. Next, define $g = I(A) + I(B)/2$ and let $T:\Delta^n \longrightarrow \Delta$ be defined by

$$T(f_1,\ldots,f_n) = f_1 g/\int f_1 g d\mu$$

(note that $1/2 \le \int f_1 g d\mu \le 1$ since $1/2 \le g \le 1$). Obviously T does not preserve unanimity and is nevertheless Externally Bayesian. However, it is neither loca, nor semi-local, nor even quasi-local!

2.5 Information maximizing and divergence minimizing pooling operators

In this section, we take a different approach to the problem of adequately describing a consensus of opinions. In the first part, we adopt the point of view that each opinion f_i contains some "information" about Θ and we look for a single representative probability distribution, $T(f_1,\ldots,f_n)$, whose expected information content will be a maximum. The pooling formula so obtained will differ according to which definition of information is elected. This approach will be seen to have the merit of providing a sensible interpretation of the constants w_i with which each opinion f_i is weighted, a question which was left unanswered by our previous attempts. Then, in the second part, we employ Kullback's (1968) concept of divergence between probability distributions to construct a class of pooling formulas which contains both the linear and the logarithmic pools as limiting cases. We begin with a short review of Shannon's definition of entropy, which is basic to the Theory of Information. For convenience, we work on Δ_0, defined in Section 2.4 to be $\{f \epsilon \Delta | f \neq 0 \ \mu\text{-a.e.}\}$.

Perhaps the most celebrated and popular measure of the amount of information contained in a probability density f on Θ is the entropy function

$$E(f) = -\int f \cdot \log(f) d\mu \quad \epsilon \ [0,\infty), \qquad (2.5.1)$$

the discrete version of which was introduced by Shannon (1948) in the context of communication engineering. The quantity E(f) measures the "uncertainty" contained in the random variable Θ (as governed by f) and thus represents, in some sense, our best knowledge of Θ. The smaller the entropy, the less uncertain is Θ and therefore the better informed one is deemed to be upon being apprised of f. Strong justification for using (2.5.1) has been supplied by way of axiomatic characterizations, though only in the discrete case. Most derivations, including those of

Faddeev (1956) and Forte (1973), are based on some version of the additivity postulate which stipulates that the information expected from two experiments equals the information expected from the first experiment plus the conditional information (entropy) of the second experiment with respect to the first. This postulate must be considered fundamental to any idea of "information."

The following expression for the entropy of one probability density f in Δ_0 with respect to another probability density g is usually known as the Kullback-Leibler Information for discriminating between f and g:

$$I(f,g) = \int f\cdot\log(g/f)d\mu. \tag{2.5.2}$$

It was defined by Shannon (1948) in the discrete case and later extended by Kullback & Leibler (1951) to the general case. The quotient $\log[g(\theta)/f(\theta)]$ may be interpreted as the "weight of evidence" (Good 1950) or the information in $\Theta = \theta$ for discriminating in favour of H_1: "the true distribution is g" versus H_0: "the true distribution is f." Alternately, the quantity (2.5.2) may be regarded as the information gain (a negative quantity here)

$$E(f) - [-\int f\cdot\log(g)d\mu]$$

incurred by using one's "best knowledge of Θ," g, to take decisions, while the true (hypothetical) underlying probability distribution governing Θ is f.

Two basic properties of the Kullback-Leibler information are embodied in

Lemma 2.5.1

Let $I:(\Delta_0)^2 \longrightarrow R$ be defined by Equation (2.5.2). If f and g represent any non-vanishing μ-densities in Δ_0, then

(i) $I(f,g) \leq 0$ always, and

(ii) $I(f,g) = 0$ iff $f=g$ μ-a.e.

Proof: This result is stated and proved as Theorem 3.1 in Chapter 2 of Kullback (1968). ■

The above properties of the Kullback-Leibler Information measure are enough to suggest a new characterization of the linear opinion pool in the following context. Let us imagine for a moment that a decision maker has collected n expert probability assessments $f_1,\ldots,f_n$ about Θ and that he is informed, knows or judges somehow that (i) one of these is the density of the "objective" probability distribution governing Θ as a random variable; and (ii) the probability that the i-th distribution, f_i, is objective is $p_i \geq 0$, $\sum_{i=1}^{n} p_i = 1$. We have already remarked in Chapter 1 that an objective distribution for

Θ may only be virtual (Θ may be observable only once, for instance), so the situation which we are describing is hypothetical. However, it is suggestive and descriptive.

If f_i were the density of the objective distribution, then according to (2.5.2), the amount of information lost due to adopting a probability distribution g instead of f_i would be

$$-I(f_i, g) = \int f_i \cdot \log(f_i/g)d\mu.$$

Averaging over the f_i's, we find that the <u>global expected information loss</u> is

$$-\sum_{i=1}^{n} p_i \cdot I(f_i, g), \qquad (2.5.3)$$

a functional depending solely on g. It would seem natural to choose g so as to <u>minimize</u> (2.5.3), i.e. pick a probability distribution which <u>minimizes the expected loss of information</u> occasioned by the need to compromise. Note that the definition of pooling operator rules out the possibility that g could be randomly chosen from the f_i's: although attractive, this selection scheme does not engender the idea of consensus. If a pooling operator $T:\Delta_0^n \longrightarrow \Delta_0$ is such that $T(f_1, \ldots, f_n) = g$

minimizes (2.5.3) whatever be $f_1,\ldots,f_n$ ($(p_1,\ldots,p_n)$ being a fixed vector of probabilities), we say that it is a <u>Kullback-Leibler Information Maximizer</u> (KLIM).

<u>Theorem 2.5.2</u>

The linear opinion pool $T(f_1,\ldots,f_n) = \sum_{i=1}^{n} w_i f_i$ is the only KLIM; moreover, $w_i = p_i$, $i=1,\ldots,n$.

<u>Proof</u>:

Call $f = \sum_{i=1}^{n} p_i f_i$. To be a KLIM, g must minimize (2.5.3) or, equivalently, maximize

$$\sum_{i=1}^{n} p_i \int f_i \cdot \log(g/f)\,d\mu = I(f,g).$$

Lemma 2.5.1 shows that $g = f$ μ-a.e. ∎

Here, we have a characterization of the linear opinion pool which does not impose a specific form on the pooling operator at the outset. Locality merely comes as a <u>consequence</u> of the definition of T. Also noteworthy is the fact that this result does not distinguish between tangible and intangible spaces.

Theorem 2.5.2 provides us with a natural interpretation of the weights, w_i, at least as they appear in the linear opinion pool. If an objective probability density, f, for Θ and objective probabilities, p_i, of $\{f_i=f\}$ exist, we have seen that w_i should equal p_i, $1 \le i \le n$. When f exists but the p_i's are unknown, it would seem natural to let w_i represent the decision maker's <u>subjective</u> probability that the i-th expert opinion is the "right one." This supports the intuitive idea that even in the absence of an objective distribution, f, the weights, w_i, should be chosen on the basis of a subjective judgement made by the decision maker concerning the accuracy of each assessor.

Winkler (1968) describes some of the most popular rules for determining the weights. All of them are based on the intuitive grounds proposed above. The most promising one, suggested by Roberts (1965), looks at likelihood ratios to compare the predictive ability of the experts; this involves the application of Bayes' Theorem to formally revise the weights after each assessment and the related observation. More simply, though, the decision maker could use the present methods to extract a consensus on the weights after having asked each expert, i, to produce a set of weights $\{w_{ij} | 1 \le j \le n\}$ on the basis of the relative importance that he would assign to the opinions of the various members of the panel, including himself. Of course, this raises further questions about the formula to be used in pooling the weights and the value (weight) to be assigned to any particular weights assessment. In principle, this process could go on for ever, except that the final consensus will generally be less sensitive to the choice of weights than to the choice of the pooling formula. "In a way,' state Mosteller & Wallace (1964, p. 264), 'this is an old story in statistics because modest changes in weights ordinarily change the output modestly."

The idea of maximizing the expected Kullback-Leibler information is not new. It has been suggested by Lindley (1956) as a sensible (but ad hoc) criterion for experimental designs "where the object of experimentation is not to reach decisions but rather to gain knowledge about the world." It is precisely the context in which this idea has been applied here: collecting expert opinions may be viewed as an experiment, and it is the stated purpose of our problem to assess the relative likelihood of the various possible states of nature, <u>not</u> to take decisions. Bernardo (1979) has shown that this maximization procedure is but another instance of the general (Bayesian) principle of maximizing the expected utility, and that it is, in some sense, the only sensible one when the object is to make inference without any specific application in mind. In Chapter 1, we drew a parallel between the problem of determining a consensus and that of estimating a quantity from a number of observations. We have seen that the linear opinion pool may be interpreted as a kind of sufficient statistic, a statistic which, according to Fisher (1934), "summarises the whole of the relevant information supplied by the sample."

It would seem natural to try to extend Theorem 2.5.2 to the so-called <u>Rényi Information measures</u>

$$I_\alpha(f,g) = (1-\alpha)^{-1}\cdot\log[\int f^\alpha g^{1-\alpha}\,d\mu], \quad 0<\alpha<1 \tag{2.5.4}$$

based on the <u>α-entropy functions</u>

$$E_\alpha(f) = (1-\alpha)^{-1}\cdot\log[\int f^\alpha\,d\mu], \quad 0<\alpha<1$$

introduced by Rényi (1961). As the reader may easily check, $I_\alpha(f,g) \to I(f,g)$ as $\alpha\to 1$ whatever be f and g in Δ; here, the restriction $\alpha < 1$ is imposed to ensure that the integral in (2.5.4) is always finite. Reasoning in the same way as before, we would like to find a possibly unique $g = P_\alpha(f_1,\ldots,f_n)$ which maximizes

$$\sum_{i=1}^{n} p_i I_\alpha(f_i,g), \tag{2.5.5}$$

the expected Rényi Information of order α, $0 < \alpha < 1$. This problem has not yet been solved for arbitrary n and α. However, a solution for the case where n=2 and $\alpha = p_1 = p_2 = 1/2$ where (2.5.5) becomes $\log\{[\int\sqrt{f_1 g}\,d\mu][\int\sqrt{f_2 g}\,d\mu]\}$ is given below.

<u>Lemma 2.5.3</u>

The quantity $[\int\sqrt{f_1 g}\,d\mu][\int\sqrt{f_2 g}\,d\mu]$ achieves a maximum when $g = H^2/\int H^2 d\mu$ μ-a.e., $H = \sqrt{f_1} + \sqrt{f_2}$.

<u>Proof</u>: Let $F_1 = \sqrt{f_1}$, $F_2 = \sqrt{f_2}$, where $H = F_1 + F_2$ and $G = \sqrt{g}$. We have

$$2\int F_1 G d\mu \int F_2 G d\mu \leq [\int F_1 G d\mu]^2 + [\int F_2 G d\mu]^2$$

and so

$$\begin{aligned} 4\int F_1 G d\mu \int F_2 G d\mu &\leq [\int (F_1+F_2) G d\mu]^2 \\ &\leq [\int (F_1+F_2)^2 d\mu]\cdot[\int G^2 d\mu] \\ &= \int H^2 d\mu. \end{aligned}$$

The second inequality is strict unless $\beta(H)^2 = \gamma g$ μ-a.e. for some $\beta,\gamma \in R$, not both zero (Rudin 1974, p. 66). If $\gamma=0$, then $\beta \neq 0$ and so $H = 0$ μ-a.e., a contradiction. Thus $\gamma \neq 0$ and $g = \beta H^2/\gamma$ with $\int g d\mu = 1$. Thus $\beta \neq 0$ and $g = H^2/\int H^2 d\mu$ μ-a.e. It so happens that the first inequality is also achieved by this particular choice of g. ∎

This partial result was obtained independently by Mr. B.J. Sharpe (private communication). It is he who pointed out that, in the setting of Example 2.5.5 below, it is possible to show that, in general, $P_a \neq (\sum_{i=1}^{n} \sqrt{f_i})^2/\int(\sum_{i=1}^{n}\sqrt{f_i})^2 d\mu$ for $n > 2$ and $w_1 = \ldots = w_n = 1/n$. The details are omitted.

We now propose to characterize the following pooling operators which we call the "normalized (weighted) means of order a:"

$$T_a(f_1,\ldots,f_n) = [\sum_{i=1}^{n} w_i f_i^{a}]^{1/a} / \int [\sum_{i=1}^{n} w_i f_i^{a}]^{1/a} d\mu, \quad 0<a<1. \qquad (2.5.6)$$

(As before, the weights w_i are non-negative and sum up to 1.)

These bear an obvious connection with the <u>weighted mean of order a</u>, $M_a(x_1,\ldots,x_n)$, of a set of n non-negative real numbers (cf. Hardy, Littlewood & Pólya 1934):

$$M_a(x_1,\ldots,x_n) = [\sum_{i=1}^{n} w_i x_i^{a}]^{1/a}.$$

These weighted means will appear again in Chapter 3, when we discuss the problem of pooling propensity functions.

The basic quantity, here, is Kullback's (1968, p. 67) notion of <u>divergence</u> between any two probability distributions, f and g:

$$\delta_a(f,g) = 1-\int f^{a} g^{1-a} d\mu, \quad 0<a<1.$$

In the case $a = 1/2$, $\delta_a(f,g)$ is equivalent to the so-called Hellinger (1909) - Kakutani (1948) - Matusita (1951) distance

$$\rho^2(f,g) = \int(\sqrt{f} - \sqrt{g})^2 d\mu,$$

a measure also used by Stein (1965) for measuring the distance between posterior distributions obtained from two different prior distributions. The function $\rho(f,g)$ is sometimes referred to as the <u>affinity</u> between f and g, after Bhattacharyya (1943).

We have the following

<u>Theorem 2.5.4</u>

The pooling operator T_a defined by Equation (2.5.6) is the only one which minimizes the expected divergence $\sum_{i=1}^{n} w_i \delta_a(f_i, g)$.

<u>Proof</u>:

Write $f = \sum_{i=1}^{n} w_i f_i^a$. By Hölder's inequality,

$$\int f g^{1-a} d\mu \le [\int f^{1/a} d\mu]^a < \infty$$

and equality is achieved only when $\beta f^{1/a} = \gamma g$ μ-a.e. for some $\beta, \gamma \in R$, not both zero (Rudin 1974, p. 66). Proceed as in the proof of Lemma 2.5.3. ∎

If a decision maker knows that each one of the n expert opinions $f_1, \ldots, f_n$ which he has collected has a corresponding probability p_i of being the "right one," then it might well seem reasonable to him to choose a consensus distribution which, on the average, will have the greatest "affinity" with the true distribution. In that case, Theorem 2.5.4 above says that T_α should be used for some $0 < \alpha < 1$. The choice of the value of α may be guided by the specific application; alternately, the decision maker might want to assess the sensitivity of his conclusions by computing a consensus for different α's.

One attractive feature of the class $\{T_\alpha\}$ is the fact that both the linear and the logarithmic pooling operators are included as limiting cases. Indeed, $T_\alpha(f_1,\ldots,f_n) \to \sum_{i=1}^{n} w_i f_i$ as $\alpha \to 1$, whatever $f_1,\ldots,f_n \in \Delta$. On the other hand, we can use L'Hospital's rule to see that

$$\begin{aligned}\log[\lim_{\alpha\to 0} (\sum_{i=1}^{n} w_i f_i^{\alpha})^{1/\alpha}] &= \lim_{\alpha\to 0} \alpha^{-1}\cdot\log[\sum_{i=1}^{n} w_i f_i^{\alpha}] \\ &= \lim_{\alpha\to 0} [\sum_{i=1}^{n} w_i f_i^{\alpha}\cdot\log(f_i)]/[\sum_{i=1}^{n} w_i f_i^{\alpha}] \\ &= \log[\prod_{i=1}^{n} f_i^{w(i)}].\end{aligned}$$

Now, it is known (Hardy, Littlewood & Pólya 1934, p. 26) that

$$[\sum_{i=1}^{n} w_i f_i^{1/k}]^{k} \le \sum_{i=1}^{n} w_i f_i$$

pointwise for all $k \geq 1$, and of course $\int \sum_{i=1}^{n} w_i f_i \, d\mu = 1$.

Therefore, we can use the Lebesgue Dominated Convergence Theorem (Sion 1968, p. 95) to conclude that

$$\lim_{k \to \infty} \int \Big[\sum_{i=1}^{n} w_i f_i^{1/k} \Big]^k d\mu = \int \lim_{k \to \infty} \Big[\sum_{i=1}^{n} w_i f_i^{1/k} \Big]^k d\mu$$

$$= \int \prod_{i=1}^{n} f_i^{w(i)} \, d\mu.$$

Consequently, $T_a \to \prod_{i=1}^{n} f_i^{w(i)} / \int \prod_{i=1}^{n} f_i^{w(i)} \, d\mu$ always as $a \to 0$. This fact may provide some indication that this quasi-local Externally Bayesian procedure is "robust" in some sense.

We conclude this section with an example borrowed from Weerahandi & Zidek (1978):

<u>Example 2.5.5</u> (Parliamentary voting procedures)
Suppose that a House of Representatives is composed of n members, each of whom has a democratic weight of 1/n when he votes. Suppose also that when a proposal is put before the House for approval, each member i tells an independent judge, Mr. Speaker say, his personal probability $0 < p_i < 1$ that passing the proposal is the right thing to do. The understanding is that this person is required to form the consensus and take a decision, approval or rejection, which is

consistent with it. Note that Mr. Speaker could have the right to vote too, as long as he does not let his personal desires influence unduly the decision ultimately made by him in his capacity as arbitrator of the group. If Mr. Speaker uses T_a ($0 < a < 1$) to establish a consensus, his arbitrator's odds in favour of passing the proposal (once he has heard every deputy) will be

$$[\sum_{i=1}^{n} p_i^a]^{1/a} / [\sum_{i=1}^{n} (1-p_i)^a]^{1/a}, \quad 0 < a \leq 1$$

$$[\prod_{i=1}^{n} p_i]^{1/n} / [\prod_{i=1}^{n} (1-p_i)]^{1/n}, \quad a = 0.$$

Thus, the best non-randomized decision rule would consist of passing the proposal if

$$(1/n)\cdot \sum_{i=1}^{n} p_i^a > (1/n)\cdot \sum_{i=1}^{n} (1-p_i)^a, \quad 0 < a \leq 1; \tag{2.5.7}$$

$$(1/n)\cdot \sum_{i=1}^{n} \log[p_i/(1-p_i)] > 0, \quad a = 0. \tag{2.5.8}$$

When $a = 1$, the procedure reduces to passing the proposal if $\bar{p} > 1/2$. When $a = 0$, the proposal will go through if, on the average, the parliamentarians' log-odds-ratios favor passage, i.e. (2.5.8) holds.

Now suppose that n_0 of the House members, $i=1,\ldots,n_0$ are against passage, and the other $n_1 = n-n_0$ are for. Further assume that $p_i = \gamma < 1/2$, $i=1,\ldots,n_0$, and $p_i = \xi > 1/2$, $i=n_0+1, \ldots,n$. Weerahandi & Zidek (1978) point out that if $\gamma = 1-\xi$, the optimal non-randomized decision rule (2.5.8) is nothing but the familiar "simple majority" voting procedure. This is also true of (2.5.7) for all $0 < \alpha \leq 1$, as the reader may easily check. Thus T_0 is not extraordinary in this respect. But the "simple majority" rule would seem well justified, at least if politicians were as certain as they appear to be in public appearances and therefore the p_i's were all essentially 0 or 1.

2.6 Discussion

The work of the previous sections has been directed toward the theoretical aspects of group probability assessment in the case where expert opinions are expressible as densities with respect to a fixed underlying measure. In particular, we have (i) proposed new arguments favouring the linear opinion pool, T_1; and (ii) characterized the logarithmic pooling operator, T_0, as the "only practical" Externally Bayesian procedure. We regard the latter result as our main contribution to this problem. Some of the following remarks will also apply in substance to the developments of Chapter 3 (they will not be repeated).

It should be clear, from the content of Section 2.5 especially, that there cannot be a unique solution to the problem of the panel of experts. This is also the conclusion reached by Bacharach (1975). Certainly, the use of either the linear or the logarithmic pool is by now well-justified, and it is interesting to think of the two as being limiting cases of an entire class of reasonable pooling formulas. This author would personally favour the logarithmic pooling operator, as he finds the axiom of coherence EB of Madansky (1964;1978) rather appealing in a Bayesian framework.

The prescription

$$\prod_{i=1}^{n} f_i^{w(i)} \Big/ \int \prod_{i=1}^{n} f_i^{w(i)} \, d\mu \qquad (2.6.1)$$

is also recommended by Weerahandi & Zidek (1978), and indirectly by Morris (1974;1977) and Winkler (1968), the latter through his natural-conjugate (N-C) approach. The N-C recipe amounts to that offered by the logarithmic opinion pool except that all probability assessments must belong to some fixed natural-conjugate family of distributions. Of course, this approach is valuable only to the extent that such a mathematical model may well approximate one's judgements. Bacharach (1973) attributes

(2.6.1) to Hammond, but he does not cite a source for the result.

As Winkler (1968) points out, the choice of a pooling operator can be influenced by practical considerations. For instance, the desire to simplify computations or the need to have an analytical expression for the consensus may well deter one from using an otherwise sensible formula. Morris' (1977) procedure, for example, entails formidable assessment problems in all but the simplest applications.

Thus, the following features of T_0 and T_1, the logarithmic and linear opinion pools respectively, would be of some relevance in the context of an actual application:

(i) <u>T_1 is generally multi-modal</u>, <u>whilst T_0 is typically uni-modal</u>

It is generally observed that the larger the differences amongst the modes of the individual probability densities f_i, the more likely it is that T_1 will produce a multi-modal distribution. The fact that T_1 may fail to identify a parameter which typifies its modes (i.e. the individual choices) might well be perceived as a fault, even if the problem does not call for a decision (cf. Weerahandi & Zidek 1981).

(ii) T_1 has a greater variance than T_0

This is not surprising in view of (i). Given a set of w_i's, the tighter distributions will automatically receive more weight under T_0 than under T_1. This is due to the multiplicative nature of the logarithmic opinion pool. For an analogy, think of the situation faced in a formal Bayesian analysis where a large amount of sample information "swamps" a relatively smaller amount of prior knowledge.

Whether a small or a large variance is more desirable will depend on the particular application one has in mind. Bernardo (1976) reports the following example: suppose that two experts gave $f_1 \simeq N(0,1)$ and $f_2 \simeq N(1+\sqrt{3}/e, e^2)$ as their respective opinions, and that θ later turned out to be 1.

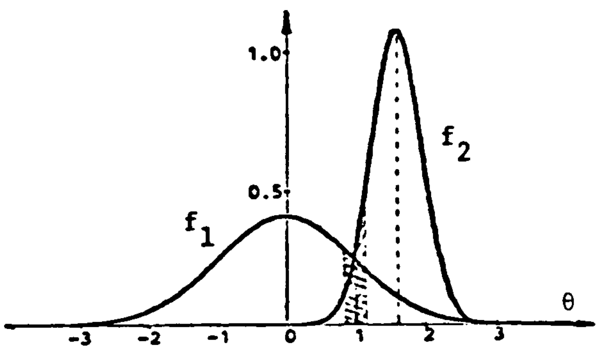

Figure 1. Two opinions with a different entropy but giving the same probability to the true value of the quantity of interest before it is revealed to be 1.

Explains Bernardo, on page 34: "In a sample survey where a loose approximation of θ may be useful, f_2 could be preferred on the grounds that it attaches a high probability to such approximate values. On the other hand, in a medical research where a small error may have fatal consequences, f_1 could be preferred on the grounds that it warns against a premature, possibly fatal decision and rather suggests that more evidence is needed." These preoccupations, however, are somewhat beyond our present concern.

(iii) <u>Calculations are easier with T_0</u>

We remark that if $f_1,\ldots,f_n$ are members of a family of exponential type determined by the same generalized density, then $T_0(f_1,\ldots,f_n)$ will be a member of the same family. For example, if f_i is a normal density with mean μ_i and variance σ_i^2, then $T_0(f_1,\ldots,f_n)$ will also be normal, with mean $\mu = \sum_{i=1}^n \alpha_i \mu_i / \sum_{i=1}^n \alpha_i$ and variance $\sigma^2 = 1/\sum_{i=1}^n \alpha_i$, where $\alpha_i = w_i/\sigma_i^2$, $1 \le i \le n$. Distributions of the form $\sum_{i=1}^n w_i f_i$ are called <u>mixtures</u> and usually are intractable, unless of course all the f_i's are the same.

Professor A.W. Marshall has raised an objection to the logarithmic opinion pool (personal communication), pointing out that it would be unsatisfactory for combining expert opinions when these opinions are based on overlapping experience or data sources. This is indeed a problem with Winkler's N-C approach, where (2.6.1) comes as a by-product of Bayes' rule. However, this is <u>not</u> a criticism of the logarithmic opinion pool which was obtained in Theorem 2.4.6. It must be interpreted instead as a criticism of the EB postulate, a logical consequence of which is the logarithmic pool. It may well be that T and the EB postulate ought not to be applied to the set $\{f_i\}$ for the reasons cited above, but instead to an alternate collection of opinions derived from the f_i's. However once the derived set is specified, EB would again lead to their logarithmic pool.

Morris (1974,1977) gets around this difficulty by encoding the degree of dependence amongst the experts as well as each expert's probability assessment ability in a so-called <u>Joint Calibration Function</u> $C:\Theta \longrightarrow [0,\infty)$. In general, this calibration function will represent the decision maker's subjective evaluation of the experts, rather than the result of the experts empirically calibrating themselves; and so the task of determining C will be rendered difficult by the need to assess the elusive dependencies between the experts' opinions.

When this is done, a "generalized logarithmic opinion pool" $M(f_1,\ldots,f_n)$ emerges for the composite prior of n experts:

$$M(f_1,\ldots,f_n) = C\prod_{i=1}^{n} f_i \Big/ \int C \prod_{i=1}^{n} f_i \, d\mu. \tag{2.6.2}$$

This reduces to (2.6.1) with $w(i) = 1, 1 \le i \le n$, if the experts are independent and calibrated (i.e. if C is a constant). This concurs with the ad hoc suggestion of Winkler (1968) that $\sum_{i=1}^{n} w_i$ should be taken in the interval [1,n] and reflect the "amount of independence" between the experts. However, M is <u>not</u> Externally Bayesian and will usually have a much smaller variance than T_0 if C is a constant (cf. remark (ii) above). Also, note that M provides us with an example of a "semi-quasi-local" pooling operator. It would be interesting to characterize all the Externally Bayesian semi-quasi-local pooling operators.

As observed by Winkler (1968) and Weerahandi & Zidek (1981), pooling operators can be used in an entirely different spirit from that which has motivated their study. Indeed, the assumption that each f_i represents a subjective probability distribution assessed by a member of a panel of experts is convenient, but not crucial to the analysis. Thus, a single individual may well choose to reflect the surmised quality of his prior knowledge in an analysis by combining this prior with "mechanical predictions" representing either ignorance or a tendency to persistence (what happened yesterday will happen today), or else derived using more complex schemes such as multiple regression. Nonetheless, the problem of determining appropriate weights remains.

Finally, a word of caution concerning the game-theoretic aspects of our problem. Throughout this chapter, we have assumed that the experts consulted by the decision maker were candid and accurate in their probability assessments. Paraphrasing Raiffa (1968), we could say that they are dedicated staff men, whose sincerity is unquestionable and who would not conspire to trick the decision maker. This is not a realistic assumption if the ultimate objective of the exercise is to make a (possibly consequential) decision and/or if the decision maker only represents the "synthetic personality" of the group. For example, one of the experts may intentionally falsify his opinion in an attempt to influence the others toward a particular consensus which is somehow advantageous to himself. Fellner (1965) discusses probabilistic "slanting" and its occurrence in group decision making. When bargaining is involved, the solutions presented here will prove unsatisfactory unless, perhaps, the members of the panel share (roughly) the same preference pattern, i.e. utility function. Unlike Weerahandi & Zidek (1981), our approach to the multi-Bayesian decision problem is through aggregation, not compromise.

III. POOLING PROPENSITIES

3.1 Motivation

Thus far, we have concentrated on the problem of reconciling judgemental probability assessments, i.e. expert opinions which are expressible as densities with respect to some natural dominating measure on a space Θ of mutually exclusive alternatives. In this chapter, we enlarge this problem and concern ourselves with what we call propensity functions, or P-functions for short. Given a space, Θ, of contemplated states of nature, a P-function is just a transformation of Θ into $(0,\infty)$. We will denote by Π the set of all P-functions on Θ.

Examples of P-functions are likelihoods, belief functions (Shafer 1976) and density functions such as those obtained from prior, vague prior, posterior, structural (Fraser 1966) and fiducial distributions. These functions, p, need not have a finite integral with respect to any particular measure; however, they share the property that $p(\theta)/p(\eta)$ represents the relative degree of support (or "propensity") expressed in favour of θ over η, θ and η being elements of Θ. This ratio $p(\theta)/p(\eta)$ may well be an odds-ratio or a likelihood-ratio, for example. In any case, the larger this quantity is, the greater is the degree of conviction in favour of θ compared to η.

To fix ideas, some specific examples will now be presented where the need will occur to pool P-functions. These applications will also serve as a motivation for the ensuing developments. In each case, a pooling operator $T:\Pi^n \longrightarrow \Pi$ is required.

<u>Example 3.1.1</u> (pooling utility functions)

When confronted with intra-group conflicts, it is sometimes necessary to take into consideration questions of utility. Typically, this will be the case if a choice or decision is to be made which will affect the members of the panel. If such a panel of experts disagrees on utility assignments for actions (or their consequences) as well as on the prior probabilities for the possible states of nature, a decision maker might choose to decompose the problem into two parts, utilities and probabilities, and proceed to extract a consensus on both matters independently. This attitude is recommended by Raiffa (1968, p. 232) despite some of its shortcomings.

Assuming that a solution to this consensus problem is sought through <u>aggregation</u>, the decision maker will have to decide on a formula for amalgamating the experts' utility functions. In that respect, the techniques of Chapter 2 are of no avail because utility functions generally do not integrate to one. On the other hand, note that any strictly positive utility function is a P-function, so that -in that case at least-pooling utility functions amounts to pooling P-functions. Furthermore, the positivity assumption can be waived if we adopt the Paretian attitude that a measure of utility is ordinal (as opposed to cardinal), and hence unique only up to a strictly monotonic increasing transformation (of course, not all such utility functions will satisfy the expected-utility principle).

Example 3.1.2 (pooling likelihoods)

In general, likelihood functions are derived from conditional probability distributions of the form P{data|true value is θ} considered as a function of θ. These conditional distributions can be subjective. For, in all but the simplest statistical applications, they are modelled in some convenient way to approximate more or less accurately the observed (but unknown) underlying distribution of the data. This mathematical modelling involves some introspection and arbitrariness on the part of the assessor, and subjective likelihoods are a reflection of that mild "interpretation" of facts and evidence.

If a problem were delicate and complicated enough, it would not be surprising that a panel of experts who were not in accord on prior probabilities for the possible states of nature were not in agreement either on the meaning and/or value of some new piece of information presented to them. At best, in those cases where the data was reliable, of good quality and related to the parameter of interest in a manner which is well understood, even critical experts would probably nearly agree on their significance and might even adopt the associated likelihood as their (common) "revised opinion." However, it is easy to conceive of situations where the new information would be so much subject to personal interpretation that its disclosure would cause the experts to disagree even further! This suggests circumstances in which expert probability assessors could be left in dissensus, even after an open and vigorous exchange of information.

One method for resolving the disagreement amongst experts about the interpretation of a set of data is to pool their associated subjective likelihood functions. When these likelihoods are non-zero everywhere on Θ (the usual case and the most interesting one), this reduces to the problem of aggregating propensities.

<u>Example 3.1.3</u> [1] (combining independent tests of hypothesis)

This is a well-known problem which has been investigated by many authors; for a general discussion and a fairly extensive bibliography on the subject, we refer to Monti & Sen (1976). The formal statistical problem may be stated as follows: given n

---

[1] We are thankful to Dr. Peter McCullagh of Imperial College (London) for suggesting this application.

independent test statistics, $J_1,\ldots,J_n$, for testing a null hypothesis $H_0:\omega\epsilon\Omega_0$ versus $H_1:\omega\epsilon\Omega_1$ ($\Omega = \Omega_0 U \Omega_1$ being a space of probability distributions), select a function, P, of $J_1,\ldots,J_n$ which is to be used as the combined test statistic. The idea behind the construction of P is that the aggregate of several tests, possibly of marginal significance individually, can lead to scientifically decisive conclusions if their results are viewed as a whole. If large values of the J_i's are considered critical for testing H_0, one common solution consists of finding a test P based on the observed <u>significance levels</u> or <u>P-values</u>, $L_i = 1-F_i(J_i)$, where $F_i(t) = P_0\{J_i<t\}$, the cumulative distribution of J_i under the null hypothesis (it is assumed that the probability distribution of each J_i is the same for all $\omega \epsilon \Omega_0$). For example, Fisher's (1932) omnibus procedure is given by $P(L_1,\ldots,L_n) = \prod_{i=1}^{n} L_i^{1/n}$ and H_0 is rejected when the observed value of P is small.

DeGroot (1973) has shown that it is possible to interpret the tail area L_i as a posterior probability or as a likelihood ratio for the acceptance of the null. Because of this, each L_i may be regarded as an individual expression of belief or P-function assessing the "likelihood" of $\Theta = \{$"H_0 is true"$\}$ (the uninteresting case where $L_i = 0$ being neglected). In that case, $P(L_1,\ldots,L_n)$ is a pooling operator acting on $\Delta^n = (0,1)^n$.

<u>Example 3.1.4</u> (The Bergson-Samuelson social-welfare function)

Economists make a distinction between what they call an individual's <u>social-welfare function</u> and his utility function (cf. Samuelson 1947, Chapter 8). A person's utility function quantifies, either cardinally or ordinally, what they prefer on the basis of their personal interests or on any other basis. On the other hand, the social-welfare function is supposed to express what the individual prefers (or, rather, would prefer) on the basis of impersonal social considerations alone. Stated differently, the social-welfare function represents the individual's "ethical" or "moral" preferences, whereas the utility function describes their "subjective" preference pattern. It is the former which the person would use if they were called upon to make a <u>moral value judgement</u>.

Mathematically, a social-welfare function is a mapping W which makes correspond a "social utility" $W(u_1,\ldots,u_n)$ to any vector $(u_1,\ldots,u_n)$ of private utilities. Provided that its domain and range are appropriately restricted, W is an instance of a P-function pooling operator. The question of defining and determining the form of a "reasonable" social-welfare function has occupied economists for some time (cf.,e.g., Harsanyi 1955). Generally speaking, there is agreement on two points: (i) W should be "local," i.e. the social utility level attached to a "prospect" should only depend on the individual utilities associated with that particular prospect; and (ii) W should be increasing in each of its arguments, the rationale being that "if you increase any agent's utility without decreasing anybody else's utility, then society is made better off."

As we shall see, these two conditions play an important role in the sequel.

Definition 3.1.5

A pooling operator $T:\Pi^n \longrightarrow \Pi$ is called local whenever there exists a function $G:(0,\infty)^n \longrightarrow (0,\infty)$ such that

$$T(p_1,\ldots,p_n)(\theta) = G(p_1(\theta),\ldots,p_n(\theta)) \qquad (3.1.1)$$

for all $\theta\epsilon\Theta$ and $p_1,\ldots,p_n \epsilon \Pi$.

Note that Equation (3.1.1) must hold everywhere, and not merely "almost everywhere." This is rendered necessary by the absence of any natural choice for a dominating measure on Θ. The following lemma gives obvious equivalent conditions for an operator to be local.

Lemma 3.1.6

A pooling operator $T:\Pi^n \longrightarrow \Pi$ is local iff

(i) $T(p_1,\ldots,p_n)(\theta) = T(p_1,\ldots,p_n)(\eta)$ whenever $p_i(\theta) = p_i(\eta)$ for all $i=1,\ldots,n$;

and

(ii) $T(p_1,\ldots,p_n)(\theta) = T(q_1,\ldots,q_n)(\theta)$ whenever $p_i(\theta) = q_i(\theta)$ for all $i=1,\ldots,n$.

Proof: This is trivial. ■

Condition (i) above could be called "consistency." If all the experts agree that states θ and η are "equiprobable" or "equally likely," then, according to Condition (i), the decision maker should attribute the same "likelihood" or "propensity" to both θ and η. To assume that the decision maker is consistent in this particular sense seems non-controversial, at least if it is believed that the assessors did not bias their judgements in the hope of gaining some strategic advantage. Condition (ii) can be regarded as a likelihood principle for P-functions, just as before (cf. page 29).

Definition 3.1.7

A pooling operator $T:\Pi^n \longrightarrow \Pi$ is said to preserve the ordering of beliefs (POB) iff

(i) $T(p_1,\ldots,p_n)(\theta) \le T(p_1,\ldots,p_n)(\eta)$ whenever $p_i(\theta) \le p_i(\eta)$ for all $i=1,\ldots,n$;

and

(ii) $T(p_1,\ldots,p_n)(\theta) < T(p_1,\ldots,p_n)(\eta)$ whenever $p_i(\theta) \le p_i(\eta)$

for all $i=1,\ldots,n$ with strict inequality for some i.

The above property corresponds to the second requirement set out by economists for the social-welfare function. Apart from its considerable intuitive appeal, we will find that it often acts as a "regularity condition" in a manner similar to the measurability assumption used in Chapter 2. Note that any POB pooling operator will satisfy Condition (i) of Lemma 3.1.6.

In Section 3.2 below, locality will be used in conjunction with four postulates of rationality in order to characterize a large class of P-function pooling operators. Amongst the requirements will appear the <u>Unanimity Principle</u>, which says that

$$p_1 = \ldots = p_n = p \quad \Rightarrow \quad T(p_1,\ldots,p_n) = p$$

whatever $p \in \Pi$ (see Axiom A below). This condition only makes sense if the scales of belief used by the different experts are <u>intercomparable</u>, i.e. if there exists an "outside" standard for the quantification of belief such as requiring that the most preferred alternative be ascribed a value of one. The difficulties associated with comparing degrees of belief are mentioned in Weerahandi & Zidek (1981); they are analogous to those which arise in the theory of utility when one attempts to compare "preferences." Although various approaches have been taken to overcome this problem in the latter context, the question remains largely unsolved (cf. Luce & Raiffa 1957; Sen 1970). In the case of degrees of belief, the difficulties which are referred to above derive from the existence of possibilities which have not yet been identified and which, therefore, are not included in Θ. Situations are conceivable where it would be natural to normalize p_i as $p_i/p_i(\theta_0)$, θ_0 being some fixed and distinguished state in Θ. In others, $p_i/\mathrm{Sup}\{p_i(\theta)\,|\,\theta\epsilon\Theta\}$ might be more natural. In others still, there might be a natural dominating measure μ on Θ with respect to which every p_i could be normalized; this, of course, is the very important special case which we discussed in Chapter 2. In general though, no particular choice seems dictated. Furthermore, certain alternatives would not always be feasible as, for example, the above-mentioned division by p_i's total mass when Θ has an infinite μ-measure.

Notwithstanding these problems associated with the intercomparability of scales of belief, we shall assume for the time being that it is reasonable to use (local) unanimity preserving pooling operators for combining P-functions. In Section 3.3, an attempt will be made at solving the more difficult problem involving the pooling of incomparable propensities.

3.2 A class of local pooling operators

We now propose certain weak and appealing conditions which, apart from locality, embrace the minimal requirements that any reasonable candidate for the role of pooling operator should satisfy (from now on, it is understood that a pooling operator acts on P-functions). These "axioms" are seen to characterize the quasi-arithmetic weighted means (defined below), a result which was proven in another context by Aczél (1948). An interesting feature of the theorem is that although all quasi-arithmetic means are continuous in their n variables, no assumption of smoothness appears in the list of axioms. When considerations relating to the scales of belief are added, a characterization of the linear and the logarithmic pooling operators are obtained.

Definition 3.2.1

A transformation $T:\Pi^n \longrightarrow \Pi$ is called a quasi-arithmetic pooling operator iff there exists a continuous and strictly increasing function $\psi:(0,\infty) \longrightarrow R$ with inverse ψ^{-1} such that

$$T(p_1,\ldots,p_n) = \psi^{-1}\left[\sum_{i=1}^{n} w_i\,\psi(p_i)\right] \qquad (3.2.1)$$

for some fixed weights $w_1,\ldots,w_n \geq 0$ with $\sum_{i=1}^{n} w_i = 1$.

Important examples of quasi-arithmetic pooling operators are the <u>linear opinion pool</u> $\sum_{i=1}^{n} w_i p_i$ $[\psi(x)=x]$, the <u>logarithmic opinion pool</u> $\prod_{i=1}^{n} p_i^{w(i)}$ $[\psi(x)=\log(x)]$ and the <u>root-mean-power pooling operator</u> $(\sum_{i=1}^{n} w_i p_i^c)^{1/c}$ $[\psi(x)=x^c, c>0]$ which includes the first one as a special case (c=1) and the second as a limiting case (c->0). The basic properties of the quantity (3.2.1) are discussed in Hardy, Littlewood & Pólya (1934) in the case where the p_i's are real numbers. Especially noteworthy is the fact that the function ψ is unique only up to an order-preserving affine transformation ax+b, a > 0. This result we record as

<u>Lemma 3.2.2</u>

Let $w_1,\ldots,w_n \geq 0$ be fixed with $\sum_{i=1}^{n} w_i = 1$ and, for j=1,2, let $G_j(\tilde{x}) = \psi_j^{-1}[\sum_{i=1}^{n} w_i \psi_j(x_i)]$ be two quasi-arithmetic weighted means such that $G_1(\tilde{x})=G_2(\tilde{x})$ whenever $x_i \in I$ for all i=1,...,n, I being some open interval in $(0,\infty)$. If there exist at least two strictly positive w_i's, then $\psi_2 = a\psi_1+b$ on I^n for some $a,b \in R$, $a>0$.

<u>Proof</u>: This result is stated and proved as Theorem 83 on page 66 in Hardy, Littlewood & Pólya (1934). ∎

We will now present four axioms which pooling operator could reasonably be required to satisfy. It will turn out that these axioms characterize the quasi-arithmetic pooling operators of Definition 3.2.1.

The central requirement is inspired by Weerahandi & Zidek's (1978) "prior-to-prior coherence" axiom and stipulates that "pooling opinions can be done sequentially and in any order." Since the ultimate objective of pooling P-functions is to construct something that can be called a combined P-function representing the beliefs of all the experts, it is plausible that the actual order in which this pooling is done should be immaterial. In their manuscript, Weerahandi & Zidek express this condition as

$$T_k(p_1,\ldots,p_k) = T_2(T_{k-1}(p_1,\ldots,p_{k-1}),p_k) \qquad (3.2.2)$$

for all $k=2,\ldots,n$, where the subscript on T indicates the dimension of its domain. Although it conveys the basic idea of sequential pooling, this formulation of the prior-to-prior coherence axiom is inadequate because it may induce an undesired functional relationship amongst the weights ascribed to the various experts. This point will be best illustrated with an example.

Suppose that we are dealing with $n=3$ experts and that expert i's opinion, p_i, has a weight of $w_i>0$, $w_1+w_2+w_3=1$. Then the relative weight of opinion p_1 with respect to p_2 is $w = w_1/(w_1+w_2)$, so $T_2(p_1,p_2)$ must assign a weight of w to its first component. However, expert 3 has a weight of w_3, which implies that the combined opinion of experts 1 and 2, p, should carry a weight of w_1+w_2. Since $T_3(p_1,p_2,p_3) = T_2(p,p_3)$, it follows that $w_1+w_2 = w = w_1/(w_1+w_2)$, from which we conclude that $w_2 = \sqrt{w_1}-w_1$ and in turn $w_3 = 1-\sqrt{w_1}$, w_1 being arbitrary in (0,1). In other words, the weight attributed to expert 1 determines the weights of both experts 2 and 3, and not only their sum!! This is clearly unacceptable.

In the above example, the difficulty arose from an ambiguity in the use of T_2 as it appears in Equation (3.2.2) when k=3. There, the "inner T_2" has the opinion of expert i as its i-th component, i=1,2, whereas the <u>joint</u> opinion of experts 1 and 2 appears in the first slot of the "outer T_2." To avoid this problem, it will be necessary to distinguish all pooling operators by keeping track of the specific weights which they ascribe to their various components, the underlying idea being that the source of an opinion is irrelevant as long as it has been properly calibrated.

<u>Change in notation</u>

When appropriate, we shall write $T_n(p_1,\ldots,p_n|w_1,\ldots,w_n)$ to denote the joint opinion of n experts $E_1,\ldots,E_n$ whose propensity functions, p_i, are to be weighted and amalgamated using formula T_n and certain weights $w_i \geq 0$, $i=1,\ldots,n$, $\sum_{i=1}^{n} w_i = 1$. If T_n is local, we shall write $T_n(p_1,\ldots,p_n|w_1,\ldots,w_n)(\theta) = G_n(p_1(\theta),\ldots,p_n(\theta)|w_1,\ldots,w_n)$.

In adopting this convention, it is understood that T_n should satisfy

$$T_n(p_1,\ldots,p_n|w_1,\ldots,w_n) = T_n(p_{T(1)},\ldots,p_{T(1)}|w_{T(1)},\ldots,w_{T(1)}) \quad (3.2.3)$$

where T is any permutation of the set $\{1,\ldots,n\}$. Strictly speaking, this requirement could be included in the following list of axioms. However, we prefer to regard it as an intrinsic property of all P-function pooling operators.

<u>Axiom A</u> (Unanimity Principle)

$T_n(p_1,\ldots,p_n|w_1,\ldots,w_n) = p$ whenever $p_1 = \ldots = p_n = p$

<u>Axiom B</u> (Preservation of the ordering of beliefs)

$$T_n(p_1,\ldots,p_n|w_1,\ldots,w_n)(\theta) \le T_n(q_1,\ldots,q_n|w_1,\ldots,w_n)(\theta)$$

whenever $p_i(\theta) \le q_i(\theta)$ for all $1 \le i \le n$, the inequality being strict if, in addition, $p_j(\theta) < q_j(\theta)$ for some $1 \le j \le n$ with $0 < w_j < 1$. When T_n is local, this is equivalent to saying that T_n is a POB pooling operator.

<u>Axiom C</u> (Prior-to-prior coherence)

For all $1 \le k \le n$,

$$T_n(p_1,\ldots,p_n|w_1,\ldots,w_n) = T_{n-k+1}(p,p_{k+1},\ldots,p_n|w,w_{k+1},\ldots,w_n)$$

where $p = T_k(p_1,\ldots,p_k|w_1/w,\ldots,w_k/w)$ and $w = \sum_{i=1}^{k} w_i$ (if $w=0$, p is an arbitrary P-function).

<u>Axiom D</u> (Monotonicity of weights)

If $w_i < w_i^*$, $w = 1 - w_i + w_i^*$ and there exists $j \ne i$ such that $w_j > 0$, then

$$T_n(p_1,\ldots,p_n|w_1,\ldots,w_n)(\theta) <$$
$$T_n(p_1,\ldots,p_n|w_1/w,\ldots,w_i^*/w,\ldots,w_n/w)(\theta)$$

provided that $p_i(\theta) = \max\{p_k(\theta) | 1 \le k \le n \ \& \ w_k > 0\}$.

When $n \geq 4$, we have the following

Theorem 3.2.3 (Aczél 1948)

Let T_n be local and suppose that there exist $T_2, \ldots, T_{n-1}$ such that Axioms A-D above be satisfied. Then T_n is a quasi-arithmetic pooling operator.

Remark 3.2.4:

The essence of the proof is contained in Aczél's (1948) paper. However, we have adapted his axioms and freed them from some obvious redundancies. For completeness, the necessary adaptation of his proof is given below.

Proof of Theorem 3.2.3:

First, we show that $T_2, \ldots, T_{n-1}$ are local whenever T_n is. For that, we define n-2 functions $G_k : (0,\infty)^k \longrightarrow (0,\infty)$, $k=2,\ldots,n-1$ by letting

$$G_k(y_1,\ldots,y_k|v_1,\ldots,v_k) = T_n(p_1,\ldots,p_k,\ldots,p_k|v_1,\ldots,v_k,0,\ldots,0)(\theta)$$

for any $p_1,\ldots,p_k \in \Pi$ with $p_i(\theta)=y_i$, $1 \leq i \leq n$. Since T_n is local, the G_k's are well-defined. Furthermore, we can use prior-to-prior coherence to see that

$$T_n(p_1,\dots,p_k,\dots,p_k|v_1,\dots,v_k,0,\dots,0)(\theta) =$$
$$T_k(p_1,\dots,p_{k-1},p|v_1,\dots,v_{k-1},v)(\theta)$$

where $p = T_{n-k+1}(p_k,\dots,p_k|v_k,0,\dots,0) = p_k$ by unanimity and $v = v_k + 0 + \dots + 0 = v_k$. Therefore,

$$T_k(p_1,\dots,p_k|v_1,\dots,v_k)(\theta) = G_k(p_1(\theta),\dots,p_k(\theta)|v_1,\dots,v_k)$$

always, i.e. T_k is local for all $k=2,\dots,n$.

Next, we define a function $\chi:[0,1]\longrightarrow[a,b]$ and verify that

$$G_2(\chi(s),\chi(t)|1-w,w) = \chi[(1-w)s+wt] \qquad (3.2.4)$$

for all $s,t,w \in [0,1]$. In fact, as is shown below, we may let $\chi(w) = G_2(a,b|1-w,w)$ for all $0\le w\le 1$, so that $\chi(0)=a$ and $\chi(1)=b$ by unanimity and prior-to-prior coherence. Using Axioms A and B, we see that $a = G_2(a,a|1-w,w) < \chi(w) < G_2(b,b|1-w,w) = b$ for all $w \in (0,1)$, and it follows from Axiom D that χ is strictly increasing in $(0,1)$.

Conjugating Equation (3.2.3) with Axioms A and C, we have successively

$$\begin{aligned}
G_2(\chi(s),\chi(t)|1-w,w) &= G_2[G_2(a,b|1-s,s),G_2(a,b|1-t,t)|1-w,w] \\
&= G_3[a,b,G_2(a,b|1-t,t)|(1-w)(1-s),(1-w)s,w] \\
&= G_4[a,b,a,b|(1-w)(1-s),(1-w)s,(1-t)w,tw] \\
&= G_4[a,a,b,b|(1-w)(1-s),(1-t)w,(1-w)s,tw] \\
&= G_3[a,b,b|1-s+ws-wt,(1-w)s,tw] \\
&= G_2[a,b|1-s+ws-wt,s-ws+wt]
\end{aligned}$$

$$= \chi(s-ws+wt) = \chi[(1-w)s+wt],$$

so that (3.2.4) holds true.

The key observation is that χ is continuous on [0,1]. For then it is surjective on [a,b] by the Intermediate Value Theorem and hence it has an inverse $\chi^{-1}:[a,b] \longrightarrow [0,1]$. This allows us to rewrite Equation (3.2.4) as

$$G_2(y,z|1-w,w) = \chi[(1-w)\chi^{-1}(y)+w\chi^{-1}(z)] \qquad (3.2.5)$$

where y and z are in [a,b] and $0 \le w \le 1$. (χ^{-1} will be the function ψ of Equation (3.2.1).)

We argue for χ's continuity by contradiction. Suppose t_0 is a point of discontinuity of χ, say to the right. Then

$$\exists y \in [a,b]\ \forall s>0 (s+t_0 \in [0,1] \Rightarrow \chi(t_0)<y<\chi(t_0+s)).$$

For such a number $y \in [a,b]$, write $y_t = G_2[\chi(t),y|1/2,1/2]$ for all $t \in [0,1]$. By Axiom B, we have

$$G_2[\chi(t),\chi(t_0)|1/2,1/2] < G_2[\chi(t),y|1/2,1/2]$$
$$= y_t < G_2[\chi(t),\chi(t_0+s)|1/2,1/2],$$

i.e., using Equation (3.2.4),

$$\chi[(t+t_0)/2] < y_t < \chi[(t+t_0+s)/2]$$

for all $s>0$ with $(t+t_0+s)/2 \in [0,1]$.

This would show that χ is discontinuous at all $(t+t_0)/2 \in [0,1]$; however, a monotone function never has more than a countable number of discontinuities (Theorem 4.30; Rudin 1976, p. 96). Hence, χ is continuous everywhere and Equation (3.2.5) obtains.

Using induction, we will now prove that

$$G_k(y_1,\ldots,y_k|v_1,\ldots,v_k) = \chi[\sum_{i=1}^{k} v_i \chi^{-1}(y_i)]$$

for all $2 \le k \le n$ and $y_i \in [a,b]$, $v_i \ge 0$, $\sum_{i=1}^{k} v_i = 1$. Indeed, we deduce from Axiom C that $G_{k+1}(y_1,\ldots,y_{k+1}|v_1,\ldots,v_{k+1})$ equals y_{k+1} if $v = \sum_{i=1}^{k} v_i = 0$ and $G_2[G_k(y_1,\ldots,y_k|v_1/v,\ldots,v_k/v),y_{k+1}|v, v_{k+1}]$ otherwise.

However, we know by hypothesis that

$$G_k(y_1,\ldots,y_k|v_1/v,\ldots,v_k/v) = \chi[\sum_{i=1}^{k}(v_i/v)\cdot\chi^{-1}(y_i)]$$

and so $G_{k+1}(y_1,\ldots,y_{k+1}|v_1,\ldots,v_{k+1})$ equals

$$G_2[\chi\{\sum_{i=1}^{k}(v_i/v)\cdot\chi^{-1}(y_i)\},y_{k+1}|v,v_{k+1}].$$

Using Equation (3.2.5) now, we find that this last quantity equals

$$\chi[v\cdot\chi^{-1}\{\chi(\sum_{i=1}^{k}(v_i/v)\cdot\chi^{-1}(y_i))\} + v_{k+1}\cdot\chi^{-1}(y_{k+1})],$$

which is nothing but $\chi[\sum_{i=1}^{k+1} v_i \chi^{-1}(y_i)]$.

To complete the proof, it remains to show that if $y_1,\ldots,y_n$ are any in $(0,\infty)$, then $\psi:(0,\infty)\longrightarrow R$ exists which is continuous, strictly increasing, and such that

$$G_n(y_1,\ldots,y_n|w_1,\ldots,w_n) = \psi^{-1}[\sum_{i=1}^{n} w_i \psi(y_i)] \qquad (3.2.6)$$

with $w_1,\ldots,w_n \ge 0$, $\sum_{i=1}^{n} w_i = 1$.

For this, we consider the nested sequence of closed intervals $I_m = [1/m+1, m+1]$ in $(0,\infty)$; we can repeat the argument above with $a=1/m+1$ and $b=m+1$ to prove the existence of a continuous and strictly increasing function $\chi_m^{-1} = \psi_m : I_m \longrightarrow [0,1]$ such that Equation (3.2.6) holds for $y_1,\ldots,y_n \in I_m$ and ψ_m instead of ψ, $m=1,2\ldots$ Since I_1 is contained in I_m and $\psi_1^{-1}[\sum_{i=1}^{n} w_i \psi_1(y_i)] = \psi_m^{-1}[\sum_{i=1}^{n} w_i \psi_m(y_i)]$ on I_1, we conclude from Lemma 3.2.2 that $\psi_m = a_m \psi_1 + b_m$ for some $a_m, b_m \in R$ with a_m strictly positive. Define $\psi^*_m = [\psi_m - b_m]/a_m$ so that ψ^*_m is an extension of ψ_1 on I_m, $m>2$; we have that ψ^*_{m+1} extends ψ^*_m from I_m to I_{m+1}, since $\psi^*_{m+1} = a\psi^*_m + b$ for some $a,b \in R$ with $a>0$, and $\psi^*_{m+1} = \psi^*_m = \psi_1$ on I_1 (i.e. $a=1,b=0$). We let $\psi = \psi^*_m$ on I_m, $m=1,2\ldots$

This completes the proof of Theorem 3.2.3. ■

Theorem 3.2.3 provides strong theoretical support for using a quasi-arithmetic pool. If a decision maker wants his pooling operator to be local, to preserve unanimity and the ordering of beliefs, as well as to satisfy the two eminently reasonable axioms of prior-to-prior coherence and monotonicity of weights, then he <u>must</u> use Formula (3.2.1), with some undetermined ψ, to pool his experts' P-functions. The requirement that $n \geq 4$ is not really limiting, because one can always throw in dummy opinions and assign them a zero weight. The only possibly controversial hypothesis, therefore, is that the various scales of belief are intercomparable; this is essential in order for the Unanimity Principle to make sense (cf. Section 3.1). As long as this assumption is valid, it seems fair to say that Aczél's conditions reflect the minimal requirements that any serious candidate to the role of P-function pooling operator would be expected to meet.

How should one choose the function ψ in Formula (3.2.1)? Clearly, there is no unique way to answer this question, at least not at this level of abstraction. Generally speaking, the criterion for selecting a "good" ψ will vary depending on the application at hand together with the intended goal. In certain circumstances, all choices of ψ will be meritorious; this will happen in the problem of pooling P-values (cf. Theorem 3.2.9 below). In other cases, however, special considerations relating to the scales of belief may induce some symmetry or invariance in the problem, and -as a result- it could seem reasonable to restrict the class of quasi-arithmetic pooling operators to those operators which are symmetric or invariant with respect to certain operations. This method is frequently used in statistical decision theory for choosing a decision rule in cases where an overall best rule does not exist. As we shall see presently, it will prove successful when searching for a Bergson-Samuelson social-welfare function, amongst others.

Examples 3.1.1 & 3.1.4 (continued)

In this example, we are concerned with finding an acceptable joint utility function, $W(u_1,...,u_n)$, which might well be a Bergson-Samuelson social-welfare function. For that purpose, we make the somewhat restrictive assumptions that the u_i's are cardinal utility functions which are bounded from below and, more importantly, which are intercomparable. The first condition implies that $u_i+c > 0$ for some $c \in R$ independent of i and so the u_i's can be simply treated as propensity functions. The second condition makes it possible to require that W preserves unanimity. However, this comparability assumption is subject to some interpretation. For instance, the u_i's might be taken to be either scale comparable or fully comparable (Sen 1970, p. 106), depending on whether we are willing to postulate that the transformed utility functions, $v_i = au_i + b_i$, are comparable for all $a > 0$ and $b_i \in R$ or only when $b_i = b$ for all $1 \le i \le n$, respectively.

Given that the utilities are fully comparable, say, Aczél's four postulates (Axioms A-D) are easily interpretable and appear to be entirely compatible with our intuitive expectations concerning the behaviour of a "reasonable" social-welfare function. By Theorem 3.2.3, the joint utility $u = W(u_1,...,u_n)$

should thus be given by

$$u = \psi^{-1}\left[\sum_{i=1}^{n} w_i \psi(u_i)\right]$$

for some continuous and strictly increasing function $\psi:(0,\infty) \longrightarrow \mathbb{R}$ and $w_1,\ldots,w_n \geq 0$, $\sum_{i=1}^{n} w_i = 1$. Furthermore, it would seem natural to demand that our pooling operator W obeys

$$W(au_1+b,\ldots,au_n+b) = a\cdot W(u_1,\ldots,u_n) + b \qquad (3.2.7)$$

for all $a>0$ and $b\in\mathbb{R}$ such that $au_i+b > 0$ on the whole domain of $u_i, i=1,\ldots,n$. This invariance property of W guarantees that no dilemma will arise from interchanging the operations of pooling and transforming the scales of belief.

As it turns out, this extra requirement is sufficient to imply a precise form for W.

<u>Proposition 3.2.5</u>

The linear opinion pool is the only quasi-arithmetic pooling operator which satisfies Equation (3.2.7) for all a and b in $(0,\infty)$.

<u>Proof</u>:

It is well known (Theorem 84; Hardy, Littlewood & Pólya 1934, p. 68) that the only quasi-arithmetic means $M(x_1,\ldots,x_n)$ which satisfy $M(ax_1,\ldots,ax_n) = a\cdot M(x_1,\ldots,x_n)$ for all $x_1,\ldots,x_n$, $a>0$ are the weighted means of order α,

$$M_\alpha(x_1,\ldots,x_n) = \begin{cases} \left[\sum_{i=1}^{n} w(i)x_i^\alpha\right]^{1/\alpha}, & \alpha\neq 0; \\ \prod_{i=1}^{n} x_i^{w(i)}, & \alpha=0, \end{cases}$$

which we have already encountered in Chapter 2 for $0<\alpha<1$. Of those, only M_1 obeys the second condition $M_\alpha(x_1+b,\ldots,x_n+b) = M_\alpha(x_1,\ldots,x_n)+b$ for all $x_1,\ldots,x_n, b > 0$.

In the language of pooling operators, this means that $W(u_1,\ldots,u_n) = \sum_{i=1}^{n} w_i u_i$ for some $w_i \geq 0$ with $\sum_{i=1}^{n} w_i = 1$ whenever W is quasi-arithmetic and Condition (3.2.7) holds. ∎

A similar argument could also be used if the u_i's were only assumed to be scale comparable. In that case, W would have the added bonus to satisfy

$$W(au_1+b_1,\ldots,au_n+b_n) = a\cdot W(u_1,\ldots,u_n) + b$$

for some $b = \sum_{i=1}^{n} w_i b_i$, $b > 0$. It may be of some interest to note that the only quasi-arithmetic pooling operators, W, which satisfy the following generalization of (3.2.7) are dictatorships:

$$W(a_1u_1+b_1,\ldots,a_nu_n+b_n) =$$
$$W(a_1,\ldots,a_n)\cdot W(u_1,\ldots,u_n) + W(b_1,\ldots,b_n).$$

If we think of a "proper" social-welfare function as one which takes account of the preference patterns of each of the individuals concerned, then this last observation could be interpreted as saying that such a function will not exist unless interpersonal comparisons of utility are possible.

Example 3.1.2 (continued)

Suppose that $p_1,\ldots,p_n \in \Pi$ represent the n opinions of a group of experts and that $\Phi_1,\ldots,\Phi_n \in \Pi$ are their respective (subjective) likelihoods for Θ given that a single data-set, $Ð = \{X_1,\ldots,X_s\}$, has been observed by all the experts. We shall assume that upon observing Ð, each expert updates his beliefs in accordance with the following rule:

$$q_i = \Phi_i \cdot p_i, \ i=1,\ldots,n, \tag{3.2.8}$$

where q_i is the opinion of the i-th expert given Ð. This formula is implied by Bayes' Theorem in the case of P-functions: the normalization constant is irrelevant as propensity functions are to be treated -and sometimes even interpreted through consideration of- betting odds. Moreover, this constant may not exist if p_i is sufficiently improper. So it is omitted and q_i is regarded as a label for the equivalence class of all constant multiples of p_i (Novick & Hall 1965, pp. 1105-1106).

The natural counterpart of an issue raised in Section 2.4 arises here: should P-functions be combined before or after the observation of the sample evidence Đ. Note that if it is decided to pool first, the discrepancy between the likelihood functions of the experts will have to be resolved before the joint P-function can be updated. The idea, here, is to pool the likelihoods, and since they are but other expressions of opinion (P-functions) from the same experts, it would seem natural to use the same pooling formula as for the priors. Moreover, it would be desirable that the operations of pooling and updating commute. An operator which does this will be called "prior-to-posterior coherent" after Weerahandi & Zidek (1978) who used it as a substitute for Madansky's term "external Bayesianity;" in this thesis, the two expressions are now vested with different meanings.

Definition 3.2.6

We say that a pooling operator $T:\Pi^n \longrightarrow \Pi$ is prior-to-posterior coherent iff

$$T(\Phi_1 p_1,\ldots,\Phi_n p_n) = T(\Phi_1,\ldots,\Phi_n)\cdot T(p_1,\ldots,p_n) \qquad (3.2.9)$$

for all $\Phi_i, p_i \in \Pi$, $i=1,\ldots,n$.

Note that this definition does not involve locality. Assuming that (3.2.8) holds, prior-to-posterior coherence is an independent criterion for selecting a pooling operator; thus, in theory at least, our search for such operators could extend to all applications $T:\Pi^n \longrightarrow \Pi$. However, arguments given above suggest that pooling formulas should be local and satisfy Aczél's four postulates, i.e. be quasi-arithmetic pooling operators. In that case, it is easy to see that Property (3.2.9) singles out the logarithmic opinion pool. In fact, we will show more:

<u>Proposition 3.2.7</u>

Let $T:\Pi^n \longrightarrow \Pi$ be a local pooling operator which preserves the ordering of beliefs (POB) and is prior-to-posterior coherent. There exist $w(1),\ldots,w(n) > 0$ such that

$$T(p_1,\ldots,p_n)(\theta) = \prod_{i=1}^{n} [p_i(\theta)]^{w(i)} \qquad (3.2.10)$$

for all $\theta \in \Theta$ and $(p_1,\ldots,p_n) \in \Pi^n$, i.e. T is a logarithmic opinion pool. Moreover, $\sum_{i=1}^{n} w(i) = 1$ whenever T preserves unanimity.

Proof:

Write $T(p_1,\ldots,p_n)(\theta) = G(p_1(\theta),\ldots,p_n(\theta))$ for all $\theta\epsilon\Theta$ and $p_1,\ldots,p_n \epsilon \Pi$. Using Equation (3.2.9), we see that $G(\tilde{x}\cdot\tilde{y}) = G(\tilde{x})\cdot G(\tilde{y})$ for all $\tilde{x}$ and $\tilde{y}$ in $(0,\infty)^n$, and it follows from Definition 3.1.7 that G is strictly increasing in each of its n variables. Appealing to Lemma 2.1.3, we conclude that $G(x_1,\ldots,x_n) = \prod_{i=1}^{n} x_i^{w(i)}$ for some $w(i)>0$, $i=1,\ldots,n$. The sum $\sum_{i=1}^{n} w(i)$ is arbitrary, unless we require that $G(x,\ldots,x)=x$ for all $x>0$, i.e. T satisfies the Unanimity Principle. ∎

To illustrate this result, suppose that a decision maker has collected some data $x_1,\ldots,x_n$ and that he regards the likelihood function $\Phi_i(\theta)$ which each of these items provides as an individual expert opinion. Formula (3.2.10) or perhaps

$$\Phi = \prod_{i=1}^{n} [\Phi_i(\theta)]^{1/n} \qquad (3.2.11)$$

might then be used to obtain the representative likelihood, Φ. Alternatively, if the sample is taken as a whole, there will be just n=1 likelihood. Both (3.2.10) and (3.2.11) would then return this likelihood in a possibly renormalized form. This result would, in general, differ from that of Equation (3.2.11).

The point, however, is that the pooling operator is not indicating, in any given context, which propensity functions should be combined, but rather it is providing a <u>means</u> of pooling such functions <u>once they have been selected</u>.

Incidentally, if the data presented in the last paragraph are obtained from independent measurements and their joint likelihood is found by first computing the joint sampling distribution and then inverting this in the usual way, the result will be $\prod_{i=1}^{n} \Phi_i$. This differs from the quantity $[\prod_{i=1}^{n} \Phi_i]^{1/n}$ which would be obtained from Equation (3.2.11). The latter is just the rescaled version of the former and, if $n \to \infty$, the Strong Law of Large Numbers implies that it converges to a constant multiple of $\exp\{I(f_\theta, f_{\theta_0})\}$, where I denotes the Kullback-Leibler discrimination measure, $\Phi_i(\theta) = f(x_i|\theta)$, $f(\cdot|\theta)$ is the sampling density if θ is the "true state of nature," and θ_0 is the true realization of the random variable Θ. If, on the other hand, the data are highly dependent, say $x_i = x_1, i=2,\ldots,n$, then Equation (3.2.11) would give, or very nearly give, the joint likelihood itself.

In Section 3.3 below, we shall address the problem of finding non-local prior-to-posterior coherent pooling operators. Let us mention in passing that a characterization of the weighted mean of order α (M_α defined above) obtains if the validity of Condition (3.2.9) is limited to those cases where there was mutual agreement <u>a priori</u> on the likelihood $\Phi = \Phi_1 = \ldots = \Phi_n$.

Proposition 3.2.8

Let $T:\Pi^n \longrightarrow \Pi$ be a quasi-arithmetic pooling operator. If T satisfies

$$T(\Phi p_1,\ldots,\Phi p_n) = \Phi \cdot T(p_1,\ldots,p_n)$$

for all Φ, $p_1,\ldots,p_n \in \Pi$, then $T = M_a$ for some $a \in (0,\infty)$.

Proof: This is because the weighted means of order $a>0$ are the only quasi-arithmetic means $M:(0,\infty)^n \longrightarrow (0,\infty)$ which are "homogeneous." See Theorem 84 on page 68 in Hardy, Littlewood & Pólya (1934). ■

Example 3.1.3 (continued)

In this case, the pooling operator $P(L_1,\ldots,L_n)$ is automatically local, because $\Theta = \{H_0\}$ is a singleton. The interpretation of Aczél's axioms causes no difficulty either. For instance, Axiom B expresses the evident requirement that a set S_1 of P-values should be more significant, as a whole, than another set S_2 if the P-values in S_1 are smaller than the corresponding P-values in S_2. As another example, the inequality $\min\{L_i \mid i=1,\ldots,n\} \le P(L_1,\ldots,L_n) \le \max\{L_i \mid i=1,\ldots,n\}$ (a consequence of Axioms A and B taken together) accounts for the fact that because the data upon which $J_1,\ldots,J_n$ are based cannot be combined directly (either because they are unavailable or incomparable due to differences in the quantitative or qualitative aspects of the various designs), we do not expect the combination test to give us more (less) confidence in H_0 than the most (least) optimistic of the observed levels L_i.

Theorem 3.2.3 suggests that we use the test statistic

$$P(L_1,\ldots,L_n) = \psi^{-1}\Big[\sum_{i=1}^{n} w_i \psi(L_i)\Big],$$

where $\psi:(0,\infty) \longrightarrow R$ would be continuous and strictly increasing on its domain. Moreover, using the fact that in general any strictly increasing continuous transformation $\chi(s)$ of a statistic S will produce the same one-sided test as S (cf. Lipták 1958, p. 176), we can restrict our attention to

$$P[\psi](L_1,\ldots,L_n) = \sum_{i=1}^{n} w_i \psi(L_i) \tag{3.2.12}$$

with suitable weights $w_i \ge 0$, $\sum_{i=1}^{n} w_i = 1$, say. This family of test statistics was first introduced by Lipták (1958) and comprises

(i) Good's (1955) weighted version of Fisher's omnibus procedure [$\psi(x)=\log(x)$]; (ii) the so-called <u>inverse normal procedure</u> [$\psi^{-1}(x)=\int I(-\infty,x)\cdot\exp\{(-1/2)(y^2+\log(2\pi)\}dy$]; and (iii) the more recent <u>logit statistic</u> of Mudholkar & George (1979) [$\psi(x)=\log(x/(1-x))$]. In the two latter cases, the domain of ψ is restricted to (0,1), but this can be justified by appealing to a different form of Theorem 3.2.3 where the P-functions would only take values in [0,1].

The following result vindicates the use of the quasi-arithmetic weighted means in this particular application.

<u>Theorem 3.2.9</u> (Lipták 1958)
Each member $P[\psi]$ of the class (3.2.12) yields a most powerful test against some specific alternative. Moreover, $P[\psi]$ is an unbiased test for the sample consisting of the P-values whenever the original test statistics $J_1,\ldots,J_n$ are.

<u>Proof</u>: A detailed proof of this theorem is contained in Lipták's paper. However, we would like to mention that the first statement is a direct consequence of the fact that any test statistic $P[\psi]$ of the form (3.2.12) satisfies Birnbaum's (1954) "condition 1." ∎

In general, the alternative against which a particular test $P[\psi]$ is admissible may be quite obscure, so this result does not constitute a strong basis for choosing one form of $P[\psi]$

over another. Moreover, the alternative is generally specified only vaguely. In his paper, Lipták proposed to circumvent this difficulty by using ψ = the inverse of the cumulative distribution for the standard normal N(0,1), i.e. the "inverse normal procedure." He claimed -without proof- that this particular choice, apart from being convenient from a computational point of view, was optimal for a large class of one-sided hypothesis testing problems, including those where the possible distributions are generated by densities belonging to the exponential family. More recently, Scholz (1981) proposed to let the P-values themselves choose the "proper" function ψ by taking that ψ (suitably standardized) which yields the largest possible value of $P[\psi]$. He describes his proposal as an application of Roy's (1953) union-intersection principle to a nonparametric setting.

Another way of comparing the $P[\psi]$'s would be to compute their Bahadur (1967) efficiency, i.e. the limiting ratio of sample sizes required by any two given test statistics of the form (3.2.12) to attain equally small significance levels. Thus, Littell & Folks (1971) showed that Fisher's method is always at least as efficient, in the Bahadur sense, as three other well-known competing methods, including Lipták's inverse normal procedure. In fact, a stronger result of the same authors has the following consequence (we are grateful to Dr. A. John Petkau for bringing this second paper of Littell & Folks to our attention):

Theorem 3.2.10 (Littell & Folks 1973)

Let $P(L_1,\ldots,L_n)$ be any test statistic which preserves the ordering of beliefs (POB) concerning the validity of the null hypothesis (Axiom B). Then P is at most as efficient, in the Bahadur sense, as Fisher's omnibus procedure.

Proof: If $P(L_1,\ldots,L_n)$ satisfies Axiom B, then the statistic $S(J_1,\ldots,J_n) = -P(L_1,\ldots,L_n)$ which rejects the null hypothesis for large values of S satisfies the condition of the theorem which appears on page 193 in Littell & Folks (1973). ∎

Remark 3.2.11

It is not too difficult to find continuous and strictly increasing functions $\psi:(0,1)\longrightarrow R$ for which the test statistic $\sum_{i=1}^{n} w(i)\psi(L_i)$ will have the same Bahadur efficiency as the corresponding weighted Fisher procedure $\prod_{i=1}^{n} L_i^{w(i)}$. For example, one could take ψ to be the inverse of the cumulative distribution of a gamma, an inverse-Gaussian or a Laplace random variable. However, Theorem 3.2.10 asserts that no test based on a statistic of the form (3.2.12) will surpass the omnibus procedure.

In summary, a set of weak and appealing conditions was developed which was shown to characterize the quasi-arithmetic pooling operators (3.2.1). It was assumed, fundamentally, that the scales of belief which were used by the various experts to express their opinions were intercomparable, so that the pooling operators could be legitimately supposed to be unanimity preserving. A locality assumption was also made for mathematical convenience. Furthermore, we saw that, depending on the application, extra invariance conditions pertaining to the scales of belief can be imposed to reduce the class of acceptable operators. However, care must be taken in imposing such conditions, as the consequent reductions may sometimes rule out "optimal" solutions. In the following section, we take a look at the more challenging problem of pooling P-functions whose scales of belief are not necessarily comparable.

3.3 Deriving the logarithmic opinion pool

It is shown in this section that the so-called general logarithmic opinion pool, Ł, is a reasonable choice of a P-function pooling operator even when degrees of belief are not intercomparable. When Θ is finite, another product formula is derived which is prior-to-posterior coherent and preserves the ordering of beliefs. At the end of the section, a parallel is drawn between our approach to pooling P-functions and Nash's (1950) solution to the multi-person cooperative decision problem.

Definition 3.3.1

The general logarithmic opinion pool is defined by

$$Ł(p_1,\ldots,p_n) = C(p_1,\ldots,p_n)\cdot \prod_{i=1}^{n} [p_i(\theta)]^{w(i)} \qquad (3.3.1)$$

for all $p_1,\ldots,p_n \in \Pi$ and $\theta\in\Theta$, where $C:\Pi^n \longrightarrow (0,\infty)$ is some unspecified function and $w(1),\ldots,w(n)$ are non-negative constants such that $\sum_{i=1}^{n} w(i) > 0$.

The operator Ł defined above is not local because the function C depends on $(p_1,\ldots,p_n)$. If $C(p_1,\ldots,p_n) = 1$ for all choices of $p_1,\ldots,p_n \in \Pi$, then Ł reduces to the logarithmic pool (3.2.10). In anticipation of the developments below, we make the following

Definition 3.3.2

The relative propensity mapping is a function $RP:\Pi^n \times\Theta^2 \longrightarrow (0,\infty)^n$ which maps any (n+2)-tuple $(p_1,\ldots,p_n,\theta,\eta)$ in $\Pi^n \times\Theta^2$ to the vector of quotients

$$RP(p_1,\ldots,p_n,\theta,\eta) = (p_1(\theta)/p_1(\eta),\ldots,p_n(\theta)/p_n(\eta)).$$

It is immediate that the application RP induces an equivalence relation on $D = \Pi^n \times\Theta^2$. If two elements of $\Pi^n \times\Theta^2$, say d_1 and d_2, are called RP-equivalent whenever $RP(d_1) = RP(d_2)$, then D may be decomposed into RP-equivalence classes obtained through RP's inverse mapping. The set $(0,\infty)^n$ may be regarded as a label set for the quotient space D/RP. As will be shown, the following property characterizes the general logarithmic opinion pool.

Definition 3.3.3

We say that a pooling operator $T:\Pi^n \longrightarrow \Pi$ is <u>relative propensity consistent</u> (RP-C) iff

$$\frac{T(p_1,\ldots,p_n)(\theta)}{T(p_1,\ldots,p_n)(\lambda)} > \frac{T(q_1,\ldots,q_n)(\eta)}{T(q_1,\ldots,q_n)(\xi)} \tag{3.3.2}$$

whenever $RP(p_1,\ldots,p_n,\theta,\lambda) > RP(q_1,\ldots,q_n,\eta,\xi)$, $p_1,\ldots,p_n$, $q_1,\ldots,q_n$ being arbitrary elements of Π and θ,η,λ,ξ belonging to Θ.

To interpret this new concept, it is useful to decompose Condition (3.3.2) into two parts, namely

(i) $$\frac{T(p_1,\ldots,p_n)(\theta)}{T(p_1,\ldots,p_n)(\lambda)} > \frac{T(p_1,\ldots,p_n)(\eta)}{T(p_1,\ldots,p_n)(\xi)}$$

whenever $RP(p_1,\ldots,p_n,\theta,\lambda) > RP(p_1,\ldots,p_n,\eta,\xi)$;

and

(ii) $$\frac{T(p_1,\ldots,p_n)(\theta)}{T(p_1,\ldots,p_n)(\lambda)} = \frac{T(q_1,\ldots,q_n)(\theta)}{T(q_1,\ldots,q_n)(\lambda)}$$

whenever $RP(p_1,\ldots,p_n,\theta,\lambda) = RP(q_1,\ldots,q_n,\theta,\lambda)$.

It is easy to check that Conditions (i) and (ii) together are equivalent to (3.3.2). The first of these conditions says that a good pooling procedure should preserve any prior consensus of the form "the odds in favour of the occurrence of θ versus λ are better than the odds for η versus ξ." In particular, note that any operator which satisfies this requirement will automatically preserve the ordering of beliefs, in the sense of Definition 3.1.7, and -by way of consequence- the "consistency" condition which appears in Lemma 3.1.6. Here again, Condition (ii) is a simplifying assumption; only this time it involves odds-ratios. It may be interpreted in the same way as the Condition (ii) which appears in Section 3.1.

We are now in a position to state and prove the principal result of this section.

Theorem 3.3.4

Suppose that Θ contains at least three distinct elements. The general logarithmic opinion pool, $\mathcal{L}$, is the only relative propensity consistent pooling operator.

Proof:

If T is any P-function pooling operator satisfying the hypotheses of the theorem, Condition (3.3.2) above has the immediate implication that the function $Q(p_1,\ldots,p_n,\theta,\eta)$ =

$T(p_1,\ldots,p_n,\theta)/T(p_1,\ldots,p_n,\eta)$ must be constant on RP-equivalence classes of D. Therefore, there exists a mapping $H:(0,\infty)^n \longrightarrow (0,\infty)$ such that $Q=H\circ RP$, the symbol $\circ$ representing as before the composition of functions.

$$\begin{tikzcd} D \arrow[r, "Q"] \arrow[d, "RP"'] & (0,\infty) \\ (0,\infty)^n \arrow[ur, dashed, "H"'] & \end{tikzcd}$$

Pick θ,η,λ, three distinct elements of Θ, and let $\tilde{x}$ and $\tilde{y}$ be two arbitrary vectors in $(0,\infty)^n$. If $p_1,\ldots,p_n \in \Pi$ are chosen so that $p_i(\theta) = x_i$, $p_i(\eta) = 1/y_i$ and $p_i(\lambda) = 1$ for all $1\le i\le n$, then $H(\tilde{x}\cdot\tilde{y}) = H(p_1(\theta)/p_1(\eta),\ldots,p_n(\theta)/p_n(\eta)) = Q(p_1,\ldots,p_n,\theta,\eta) = Q(p_1,\ldots,p_n,\theta,\lambda)\cdot Q(p_1,\ldots,p_n,\lambda,\eta) = H(p_1(\theta),\ldots,p_n(\theta))\cdot H(1/p_1(\eta),\ldots,1/p_n(\eta)) = H(\tilde{x})\cdot H(\tilde{y})$. It also follows from the RP-C condition that H is non-decreasing in each of its n variables, so that $H(\tilde{x}) = \prod_{i=1}^{n} x_i^{w(i)}$ with some fixed numbers $w(1),\ldots,w(n) > 0$ by an application of Lemma 2.1.3 (here, POB plays its role as a regularity condition).

Consequently, $Q(p_1,\ldots,p_n,\theta,\eta) = \prod_{i=1}^{n} [p_i(\theta)/p_i(\eta)]^{w(i)}$ for all

$p_1, \ldots, p_n \in \Pi$ and $\theta, \eta \in \Theta$, i.e. we have shown that the function

$$C(p_1, \ldots, p_n)(\theta) = T(p_1, \ldots, p_n)(\theta) / \prod_{i=1}^{n} [p_i(\theta)]^{w(i)}$$

is independent of θ! ∎

This theorem is not true if Θ contains exactly two elements, as the following counter-example shows.

<u>Example 3.3.5</u>

Let $\Theta = \{\theta, \eta\}$ and define $T: \Pi^n \longrightarrow \Pi$ by $T(p_1, \ldots, p_n) = (p_1 + p_2)/2$.

Then T is relative propensity consistent and even unanimity preserving, but clearly $T \neq Ł$.

As before, the problem of choosing the weights w(i) remains and is not addressed here. Note also that the function C in Equation (3.3.1) is undetermined, except for the trivial requirement $C(p, \ldots, p) = 1$ for all $p \in \Pi$ if Ł satisfies the Unanimity Principle (and the experts' scales of belief are intercomparable). At the moment, it is not clear to us what role this function plays or even how it could be interpreted. If we insist that Ł should be prior-to-posterior coherent in the sense of Definition 3.2.6, it is necessary to have $\sum_{i=1}^{n} w(i) = 1$ and also $C(p_1, \ldots, p_n) \cdot C(q_1, \ldots, q_n) = C(p_1 \cdot q_1, \ldots, p_n \cdot q_n)$ for all $(p_1, \ldots, p_n), (q_1, \ldots, q_n) \in \Pi^n$; but even that requirement is not strong enough to completely determine C.

The following partial result gives still another indication of the large variety of pooling operators which is encompassed by the term "non-local."

Proposition 3.3.6

Let $\Theta = \{\theta_1,\ldots,\theta_m\}$ be finite, and assume that $T:\Pi^n \longrightarrow \Pi$ is a POB pooling operator which is prior-to-posterior coherent. There exists a set $\{w(i,j,k) \mid 1\le i\le n, 1\le j,k\le m\}$ of positive constants such that

$$T(p_1,\ldots,p_n)(\theta_k) = \prod_{j=1}^{m} \prod_{i=1}^{n} [p_i(\theta_j)]^{w(i,j,k)}$$

for all $p_1,\ldots,p_n \in \Pi$ and $k\in\{1,\ldots,m\}$.

Proof:

Fix $\theta_k \in \Theta$ and consider $H(p_1,\ldots,p_n) = T(p_1,\ldots,p_n)(\theta_k)$ as a group homomorphism between Π^n and $(0,\infty)$ with multiplication. Then $H(p_1,\ldots,p_n) = \prod_{i=1}^{n} H_i(p_i)$ where $H_i:\Pi \longrightarrow (0,\infty)$ is defined by $H_i(p_i) = H(1,\ldots,p_i,\ldots,1)$ for each $i=1,\ldots,n$, and we have $H_i(p\cdot q) = H_i(p)\cdot H_i(q)$ whenever p and q belong to Π. Each H_i can be decomposed further as $H_i(p) = \prod_{j=1}^{m} H_{ij}(p(\theta_j))$ where H_{ij} :

$(0,\infty) \longrightarrow (0,\infty)$ is defined by $H_{ij}(x) = H_i(\tilde{x})$, where $\tilde{x} \in \Pi$ is that P-function whose value at θ_j is x and 1 otherwise. Now, H_{ij} is a homomorphism on $(0,\infty)$, and it is non-decreasing because T preserves the ordering of beliefs. Using Lemma 2.1.3, it follows that $H_{ij}(x) = x^{w(i,j,k)}$ for some $w(i,j,k) > 0$, with k indicating a possible dependence on θ_k. Combining all these facts, we obtain the desired conclusion. ■

To conclude this section, we would like to draw a parallel between the general logarithmic opinion pool and Nash's (1950) solution to the so-called "bargaining problem." The two are related in a manner which will now be described.

In general, given an action space A and a space of randomized decision rules D*, let

$$u_i(\zeta) = \int\int u_i(a,\theta)\zeta(da)b_i(\theta)\nu(d\theta),$$

where u_i denotes the i-th player's utility function and b_i is a prior or posterior distribution, whichever is appropriate. Then Nash's axioms imply that the solutions, ζ^*, are those which maximize the symmetric product $\prod_{i=1}^{n} [u_i(\zeta)-c_i]^{1/n}$ where c_i

denotes the i-th player's status quo point, i.e. the amount in utility which he will have in the event that the group fails to agree on a choice for ζ. This maximization is subject to $u_i(\zeta) > c_i$ for all $i=1,\ldots,n$ (see Weerahandi & Zidek 1981 for further details).

Now suppose $\Theta = \{\theta_1,\ldots,\theta_m\}$ is finite, as in Proposition 3.3.6. Assume further that $u_i(a,\theta) = \delta(a,\theta) + c_i / \sum_{j=1}^{m} b_i(\theta_j)$ where $\delta(a,\theta)$ is the Kronecker delta function. Then the Nash solution maximizes

$$\prod_{i=1}^{n} [\int\int \delta(a,\theta)\zeta(da) b_i(\theta)\nu(d\theta)]^{1/n} = \prod_{i=1}^{n} [\int b_i(a)\zeta(da)]^{1/n},$$

where ν denotes the usual counting measure. Observe that if ζ is restricted to be nonrandomized, the optimal choice of a is the θ which maximizes $\prod_{i=1}^{n} [b_i(\theta)]^{1/n}$, essentially the quantity which would be obtained from Equation (3.3.1) with $w(1) = \ldots = w(n) = 1/n$. This observation lends some additional support to the general logarithmic pooling recipe.

Clearly, a similar argument could be found for Stone's linear pooling operator by appealing to the work of Bacharach (1975) which in turn relies on an unpublished contribution of

Madansky. This work includes a theorem which shows that the optimal decision rule maximizes $\sum_{i=1}^{n} w_i u_i(\zeta)$. Indeed, the sort of reduction which is sketched in the last paragraph would yield the linear opinion pool. It should be noted, however, that Bacharach's result implicitly assumes the intercomparability of utilities. This implicit hypothesis arises when a theorem of Blackwell & Girshick (1954) is invoked in proving the asserted conclusion. The Blackwell-Girshick result deals with the classical decision problem where the i's represent different states of nature, <u>not</u> different players. In that situation, there is only one player and presumably he would have no difficulty comparing his own preferences and hence deducing the utility functions for the different states of nature, i.

In certain decision or estimation problems, a propensity function may be used to find the θ in Θ for which there is the maximum joint propensity (MJP). If, for example, $b_i(\theta) = \exp\{\theta u_i - \Lambda(\theta)\}$, the θ of MJP is the unique solution of $\Lambda'(\theta) = \bar{\mu} = \sum_{i=1}^{n} \alpha_i \mu_i$. This has the curious consequence pointed out in Weerahandi & Zidek (1978) that if the μ_i's are widely separated, $\bar{\mu}$ may well have a very low propensity as measured by the individual P-functions, b_i. The corresponding difficulty with the linear opinion pool is that the joint propensity function in this case is multi-modal. In place of a single representative μ (say $\bar{\mu}$) of low propensity relative to each b_i, a family of nonrepresentative μ's (approximately the μ_i's themselves) is obtained, each of high propensity relative to exactly one of these b_i's.

By extending the domain of the pooling operator from Θ to D* in the manner advocated by Weerahandi & Zidek (1978), the anomalies described in the previous paragraph can easily be circumvented. In this extension, which is suggested by the Nash theory sketched above, $b_i(\theta)$ is replaced by $b_i(\zeta) = \int b_i(\theta)\zeta(d\theta)$ and the θ of MJP is replaced by the ζ of MJP; in the situation we have just described, this would lead to a randomized choice amongst, approximately, the widely separated μ_i's. The various pooling operators suggested in this chapter would remain unchanged under this extension of domain. In fact, their derivation would go through in exactly the same way, since the $b_i(\zeta)$'s are P-functions too!

Finally, it should be added that we have not attempted to extend our definition of propensity function to comprise those whose range includes zero. Most derivations of this chapter would run into difficulty in this case. Again, the problem we face in this situation is not unlike that encountered in conventional Bayesian analysis, when the prior and the likelihood functions have disjoint supports and some improvisation is in order.

IV. SUGGESTIONS FOR FURTHER RESEARCH

The present work has reinforced, if anything, the view that there is no "one best way" to aggregate expressions of belief. Along with Savage (1971), one could say that "what to do when doctors disagree has always been, and will always be a quandary." However, there is great potential value in the idea of choosing amongst the infinite number of synthetic experts that constitute the convex closure of a panel; witness the various studies (cf., e.g., Sanders 1963; Winkler 1971; Shapiro et al. 1977,1979) in which composite distributions have been observed to predict with greater accuracy than most individual experts. Granted, the evidence is still mainly empirical and remains to be probed with sound analytical models; but, says Hogarth (1977, p. 241), "there would seem to be little doubt that the general results concerning the reliability and validity of average judgments in the form of point estimates... will carry over to probability distributions." The development of such models is one urgent task that lies ahead of us.

In the paper of Savage which is cited above, the author argues -albeit indirectly- that the chronic lack of dependable techniques of communication from which our society suffers makes the aggregation of opinions difficult. This difficulty is reduced at least in the case which we described where it is specified at the outset that the purpose of the decision maker is "to learn more" or perhaps to make a forecast, but not to take decisions. A lack of candor or ineffective communication, gross exaggeration or excessive deference, for example, are more likely to be witnessed in situations where a group of experts finds that they are competing with one another for some strategic advantage. Moreover, even in decision-oriented tasks, one would expect such solidly established rules as the linear or the logarithmic opinion pools to show some robustness to "reasonable" departures from the assumption of nearly identical utility functions for the experts. Nevertheless, we agree with Savage that the need is there to devise appropriate methods of communication which will help a panel to honestly share all relevant factual information and make it possible for its members to assist one another in thinking their beliefs through thoroughly.

To a large degree, the problem of choosing an appropriate scheme for weighing individual opinions also remains unsolved. Efforts in this direction have been made by Roberts (1965) and Winkler (1971), inter alia. However, no wholly satisfactory solution to this problem is yet in sight. In fact, the issue would seem to have been made more complicated in the light of the results of Section 3.3 which reveal that the weights could vary with θ, the parameter of interest. The only firm recommendation which can currently be made to decision makers seeking to find a consensus for a panel of experts would seem to be that they should use <u>sensitivity analysis</u> to identify the crucial aspects of the weight allocation task.

In a recent technical report, Zidek (1982) introduced two new criteria for assessing group decision procedures. One is based on the idea of subsampling the group and it is found that amongst the proposed solution concepts only Nash's solution is optimal under "subsampling." The other assumes that the group is itself a sample from a superpopulation, and this yields an analogue of Wald's theory where the elicitation of the priors becomes part of the experimental process. It will be interesting to see what these ideas will yield when they are applied in the present context.

Finally, we list a number of unsettled technical issues which were raised in the course of the discussion:

1. We proved in Section 2.3 that if T is any dogma preserving semi-local pooling operator whose underlying function "G" on $[0,\infty)^n$ is continuous, then T is also local and hence a linear opinion pool. As noted in the second paragraph following Proposition 2.3.6, this condition on G seems rather artificial and, considering the way in which it was used in our proof, could conceivably be weakened, if not removed altogether. How?

2. In Theorem 2.4.6, it was seen that the logarithmic opinion pool is the only available Externally Bayesian quasi-local pooling operator when (Θ,μ) contains non-negligible sets of arbitrary small measure (Assumption 2.4.5). Can this

somewhat irritating regularity condition be weakened or, better still, eradicated? It would be particularly important to find out whether there are or are not any other quasi-local Externally Bayesian procedures than the logarithmic pool (allowing for negative weights w_i as long as $\sum_{i=1}^{n} w_i = 1$) when Θ is finite and μ is a counting-type measure.

3. The problem of determining which $g \in \Delta_0$ maximizes the product

$$P = \prod_{i=1}^{n} [\int f_i^{\alpha} g^{1-\alpha} d\mu]^{w(i)}$$

arose in Section 2.5 when we were trying to determine which μ-density g optimized the expected Rényi Information $\sum_{i=1}^{n} w_i I_{\alpha}(f_i, g)$ of order α, $\alpha \in (0,1)$. This would seem to be a hard problem, but it may nevertheless be possible to solve it analytically.

4. It was suggested in Chapter 3 that the rule (3.2.8) could be viewed as an analogue of Bayes' formula for updating P-functions. However, arguments in favour of its use could be best developed within the framework of an axiomatic theory of propensity functions.

These questions remain for future consideration.

REFERENCES

ACZEL, János (1948)
Un problème de M.L. Fejér sur la construction de Leibniz.
Bull. Sci. Math.,72,39-45

ACZEL, János (1966)
Lectures on functional equations and their applications.
New York: Academic Press

ACZEL, János & DAROCZY, Zoltán (1975)
On measures of information and their characterizations.
New York: Academic Press

BACHARACH, Michael (1973)
Bayesian dialogues.
Unpublished manuscript - Christ Church, Oxford

BACHARACH, Michael (1975)
Group decisions in the face of differences of opinion.
Manag. Sci.,22,182-191

BAHADUR, Raghu Raj (1967)
Rates of convergence of estimates and test statistics.
Ann. Math. Statist.,38,303-324

BERGER, Roger L. (1981)
A necessary and sufficient condition for reaching a consensus using DeGroot's method.
J. Amer. Statist. Assoc.,76,415-418

BERNARDO, José M. (1976)
The use of information in the design and analysis of scientific experimentation.
Ph. D. Thesis, University of London

BERNARDO, José M. (1979)
Expected information as expected utility.
Ann. Statist.,7,686-690

BHATTACHARYYA, A. (1943)
On a measure of divergence between two statistical populations defined by their probability distributions.
Bull. Cal. Math. Soc., 35, 99-109

BICKEL, Peter J. & DOKSUM, Kjell A. (1977)
Mathematical statistics: basic ideas and selected topics.
San Francisco: Holden-Day

BIRNBAUM, Allan (1954)
Combining independent tests of significance.
J. Amer. Statist. Assoc., 49, 559-574

BLACKWELL, David H. & GIRSHICK, Meyer A. (1954)
Theory of games and statistical decisions.
New York: John Wiley

DALKEY, Norman C. (1975)
Towards a theory of group estimation.
The Delphi method: techniques and applications.
(H. A. Linstone & M. Turoff, editors)
Reading, Mass: Addison-Wesley

DAWID, A. Philip (1982)
The well-calibrated Bayesian (with discussion).
J. Amer. Statist Assoc., 77, 605-613

DAWID, A.P., STONE, M. & ZIDEK, J.V. (1973)
Marginalization paradoxes in bayesian and structural inference (with discussion).
J. R. Statist. Soc. B, 35, 189-233

de FINETTI, Bruno (1937)
La prévision: ses lois logiques, ses sources subjectives.
Ann. Inst. Henri Poincaré, 7, 1-68

DeGROOT, Morris H. (1973)
Doing what comes naturally: interpreting a tail area as a posterior probability or as a likelihood ratio.
J. Amer. Statist. Assoc., 68, 966-969

DeGROOT, Morris H. (1974)
Reaching a consensus.
J. Amer. Statist. Assoc.,69,118-121

EISENBERG, Edmund & GALE, David (1959)
Consensus of subjective probabilities: The pari-mutuel method.
Ann. Math. Statist.,30,165-168

FADDEEV, Dmitrii K. (1956)
On the concept of entropy of a finite probabilistic scheme (in Russian).
Uspehi Mat. Nauk,11 (1),227-231

FELLNER, William (1965)
Probability and profit.
Homewood, Ill: Irwin

FINE, Terrence L. (1973)
Theories of probability.
New York: Academic Press

FISHER, Sir Ronald A. (1921)
On the mathematical foundations of theoretical statistics.
Philos. Trans. Roy. Soc. London A,222,309-368

FISHER, Sir Ronald A. (1932)
Statistical methods for research workers (4th edition).
London: Oliver & Boyd

FORTE, Bruno (1975)
Why Shannon's entropy.
Symp. Math. Vol. XV,137-152
New York: Academic Press

FRASER, Donald A.S. (1966)
Structural probability and a generalization.
Biometrika,53,1-9

FRENCH, Simon (1981)
Consensus of opinion.
Europ. J. Oper. Res.,7,332-340

FRENCH, Simon (1982)
On the axiomatisation of subjective probabilities.
Theory and Decision,14,19-33

GENEST, Christian (1982)
External bayesianity: An impossibility theorem.
Tech. Rep. 82-8
Instit. Appl. Math. Statist., UBC

GENEST, C., WEERAHANDI, S. & ZIDEK, J.V. (1982)
The logarithmic opinion pool.
Tech. Rep. 82-7
Instit. Appl. Math. Statist., UBC

GOOD, Irving John (1950)
Probability and the weighing of evidence.
New York: Griffin, London & Hafner

GOOD, Irving John (1955)
On the weighted combination of significance tests.
J. Roy. Statist. Soc.,17,264-265

HARDY, G.H., LITTLEWOOD, J.E. & POLYA, G. (1934)
Inequalities.
Cambridge: Cambridge University Press

HARSANYI, John C. (1955)
Cardinal welfare, individualistic ethics and interpersonal comparisons of utility.
Journal of Political Economy,63,309-321

HELLINGER, Ernst (1909)
Neue Begründung der Theorie quadratischer Formen von unendlichen Veränderlichen.
J. Reine Angew. Math.,136,210-271

HOGARTH, Robin M. (1975)
Cognitive processes and the assessment of subjective probability distributions (with discussion).
J. Amer. Statist. Assoc.,70,271-294

HOGARTH, Robert M. (1977)
Methods for aggregating opinions.
Decision Making and Change in Human Affairs.
(H. Jungermann & G. deZeeuw, editors)
Dordrecht, Holland: D. Reidel Publishing Co.

JEFFREYS, Sir Harold (1967)
Theory of probability (3rd revised edition).
Oxford University Press

KAKUTANI, Shizuo (1948)
On equivalence of infinite product measures.
Ann. Math.,49,214-224

KOOPMAN, Bernard O. (1940)
The bases of probability.
Bull. Amer. Math. Soc.,46,763-774

KULLBACK, Solomon (1968)
Information theory and statistics.
New York: Dover Publications

KULLBACK, Solomon & LEIBLER, Richard A. (1951)
On information and sufficiency.
Ann. Math. Statist.,22,79-86

LINDLEY, Dennis V. (1956)
On a measure of the information provided by an experiment.
Ann. Math. Statist.,27,986-1005

LINDLEY, Dennis V. (1961)
The use of prior probability distributions in statistical inference and decisions.
Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol.I,453-468
Berkeley: University of California Press

LINDLEY, Dennis V. (1982)
Scoring rules and the inevitability of probability (with discussion).
Int. Statist. Review,50,1-26

LINDLEY, D.V., TVERSKY, A. & BROWN, R.V. (1979)
On the reconciliation of probability assessments (with discussion).
J. R. Statist. Assoc A,142,146-180

LIPTAK, Tamás (1958)
On the combination of independent tests.
Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei,3,171-197

LITTELL, Ramon C. & FOLKS, J. Leroy (1971)
Asymptotic optimality of Fisher's method of combining independent tests.
J. Amer. Statist. Assoc.,66,802-806

LITTELL, Ramon C. & FOLKS, J. Leroy (1973)
Asymptotic optimality of Fisher's method of combining independent tests II.
J. Amer. Statist. Assoc.,68,193-194

LUCE, R. Duncan & RAIFFA, Howard (1957)
Games and decisions.
New York: John Wiley

MADANSKY, Albert (1964)
Externally bayesian groups.
RAND Memo RM-4141-PR

MADANSKY, Albert (1978)
Externally bayesian groups.
Unpublished manuscript - University of Chicago

MARSHALL, Albert W. & OLKIN, Ingram (1979)
Inequalities: theory of majorization and its applications.
New York: Academic Press

MATUSITA, Kameo (1951)
On the theory of statistical decision functions.
Ann. Inst. Math. Statist.,3,17-35

McCONWAY, Kevin J. (1981)
Marginalization and linear opinion pools.
J. Amer. Statist. Assoc.,76,410-414

MONTI, Katherine L. & SEN, Pranab K. (1976)
The locally optimal combination of independent test statistics.
J. Amer. Statist. Assoc.,71,903-911

MORRIS, Peter A. (1974)
Decision analysis expert use.
Manag. Sci.,20,1233-1241

MORRIS, Peter A. (1977)
Combining expert judgments: a bayesian approach.
Manag. Sci.,23,679-693

MOSTELLER, Frederick & WALLACE, David L. (1964)
Inference and disputed autorship: the federalist.
Reading, Mass.: Addison-Wesley

MUDHOLKAR, Govind S. & GEORGE, E. Olusegun (1979)
The logit statistic for combining probabilities - an overview.
Optimizing methods in statistics.
(J.S. Rustagi, editor)
New York: Academic Press

NASH, John F. Jr. (1950)
The bargaining problem.
Econometrica,18,155-162

NORVIG, Torsten (1967)
Consensus of subjective probabilities: a convergence theorem.
Ann. Math. Statist.,38,221-225

NOVICK, Melvin R. & HALL, W. Jackson (1965)
A bayesian indifference procedure.
J. Amer. Statist. Assoc.,60,1104-1117

PRESS, James S. (1978)
Qualitative controlled feedback for forming group judgements and making decisions.
J. Amer. Statist. Assoc.,73,526-535

RAIFFA, Howard (1968)
Decision analysis: introductory lectures on choices under uncertainty.
Reading, Mass: Addison-Wesley

RENYI, Alfréd (1961)
On measures of entropy and information.
Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. I,453-468
Berkeley: University of California Press

ROBERTS, Harry V. (1965)
Probabilistic prediction.
J. Amer. Statist. Assoc.,60,50-62

ROY, S.N. (1953)
On a heuristic method of test construction and its use in multivariate analysis.
Ann. Math. Statist.,24,220-238

ROYDEN, Halsey L. (1968)
Real analysis.
New York: Macmillan

RUDIN, Walter (1974)
Real and complex analysis.
New York: McGraw-Hill

RUDIN, Walter (1976)
Principles of mathematical analysis (3rd edition).
New York: McGraw-Hill

SAMUELSON, Paul Anthony (1947)
Foundations of economic analysis.
Harvard University Press

SANDERS, Frederick (1963)
On subjective probability forecasting.
Journal of Applied Meteorology, 2,191-201

SAVAGE, Leonard J. (1971)
Elicitation of personal probabilities and expectations.
J. Amer. Statist. Assoc.,66,783-801

SCHOLZ, Fritz W. (1981)
Combining independent P-values.
Tech. Rep. 11
Department of Statistics, University of Washington (Seattle)

SEN, Amartya K. (1970)
Collective choice and social welfare.
San Francisco: Holden-Day

SHAFER, Glenn (1976)
A mathematical theory of evidence.
Princeton, N.J.: Princeton University Press

SHANNON, Claude E. (1948)
A mathematical theory of communication.
Bell System Technical Journal, 27,379-423 and 623-656

SHAPIRO, Alan R. (1977)
The evaluation of clinical predictions: a method and initial applications.
New Engl. J. Med.,296,1509-1514

SHAPIRO, A.R., DANNENBERG, A.L. & FRIES, J.F. (1979)
Enhancement of clinical predictive ability by computer consultation.
Meth. Inform. Med.,18,10-14

SION, Maurice (1968)
Introduction to the methods of real analysis.
Toronto: Holt, Rinehart & Winston

SLOVIC, P., FISCHHOFF B. & LICHTENSTEIN, S. (1977)
Behavioral decision theory.
Ann. Rev. Psychol.,28,1-39

STAEL VON HOLSTEIN, Carl-Axel S. (1970)
Assessment and evaluation of subjective probability distributions.
Stockholm: Economics Research Institute at the Stockholm School of Economics

STEIN, Charles M. (1965)
Approximations of improper prior measures by prior probability measures.
Bernoulli, Bayes, Laplace Anniversary Volume
(J. Neyman & L. LeCam, editors)
New York: Springer-Verlag

STONE, Mervyn (1961)
The linear opinion pool.
Ann. Math. Statist.,32,1339-1342

STONE, Mervyn (1976)
Strong inconsistency from uniform priors (with discussion).
J. Amer. Statist. Assoc.,71,114-125

TVERSKY, Amos & KAHNEMAN, Daniel (1974)
Judgment under uncertainty: heuristics and biases.
Science,185,1124-1131

WAGNER, Carl (1982)
Allocation, Lehrer models, and the consensus of probabilities.
Theory and decision,14,207-220

WEERAHANDI, Samaradasa & ZIDEK, James V. (1978)
Pooling prior distributions.
Tech. Rep. 78-34
Instit. Appl. Math. Statist., UBC

WEERAHANDI, Samaradasa & ZIDEK, James V. (1981)
Multi-bayesian statistical decision theory.
J. R. Statist. Soc. A,144,85-93

WINKLER, Robert L. (1967)
The assessment of prior distributions in bayesian analysis.
J. Amer. Statist. Assoc.,62,776-800

WINKLER, Robert L. (1968)
The consensus of subjective probability distributions.
Manag. Sci.,15,B61-B75

WINKLER, Robert L. (1971)
Probabilistic prediction: some experimental results.
J. Amer. Statist. Assoc.,66,675-685

ZELLNER, Arnold (1977)
Maximal data information prior distributions.
New Developments in the Application of Bayesian Analysis
(A. Aykaç and C. Brumat, editors)
Amsterdam: North-Holland

ZIDEK, James Victor (1982)
Aspects of multibayesian theory.
Tech. Rep. 82-11
Instit. Appl. Math. Statist., UBC