Generalising the Maximum Entropy Inference Process to the Aggregation of Probabilistic Beliefs
Version 5

George Wilmers

November 1, 2011

1 Introduction

The present paper stems from a desire to combine ideas arising from two historically different axiomatic frameworks of probabilistic reasoning, each having its own traditions, into a single broader axiomatic framework, capable of providing general new insights into the nature of probabilistic inference in a multiagent context. It is an extended elaboration and clarification of an approach previously announced in [38].

In this introduction we describe briefly the background context to the conceptual framework which we will introduce. In Section 2 we present a set of natural principles to be satisfied by any general method of aggregating the partially defined probabilistic beliefs of several agents into a single probabilistic belief function. We will call such a general method of aggregation a social inference process. We discuss in this context what appears to be a novel principle which we call the Collegial Principle, which we claim that any social inference process should satisfy. In Section 3 we define a particular social inference process, the Social Entropy Process (abbreviated to **SEP**), which satisfies the principles formulated earlier. **SEP** has a natural justification in terms of information theory, and is closely related to both the maximum entropy inference process and the notion of minimum cross-entropy. Indeed one of the main the principal aims of the present work is to argue that **SEP** should be regarded as the natural extension of the maximum entropy inference process to the multiagent context. Finally in Section 4 we give an alternative characterisation of **SEP** which has a natural heuristic justification.

By way of comparison, for any appropriate set of partial probabilistic beliefs of an *isolated* individual the well-known maximum entropy inference process, **ME**, chooses a probabilistic belief function consistent with those beliefs. We conjecture that **SEP** is the only "natural" social inference process which extends **ME** to the multiagent case, always under the assumption that no additional information is available concerning the expertise or other properties of the individual agents[1].

In order to fix notation let $S = \{\alpha_1, \alpha_2, \ldots \alpha_J\}$ denote some fixed finite set of mutually exclusive and exhaustive atomic events, or, as we prefer to think of them in a logical framework, atoms of some finite Boolean algebra of propositions. We shall refer to $S = \{\alpha_1, \alpha_2, \ldots \alpha_J\}$ as *atoms.* A probability function $\boldsymbol{w}$ on S is a function $\boldsymbol{w} : S \to [0,1]$ such that $\Sigma_{j=1}^{J} \boldsymbol{w}(\alpha_j) = 1$. Slightly abusing notation we will identify $\boldsymbol{w}$ with the vector of values $< w_1 \ldots w_J >$ where w_j denotes $\boldsymbol{w}(\alpha_j)$ for $j = 1 \ldots J$. If such a $\boldsymbol{w}$ represents the subjective belief of an individual **A** in the outcomes of S we refer to $\boldsymbol{w}$ as **A**'s *belief function.* All other more complex events considered are equivalent to disjunctions of the α_j and are represented by the Greek letters θ, ϕ, ψ etc. A probability function $\boldsymbol{w}$ is assumed to extend so as to take values on complex events in the standard way, i.e. for any θ

$$\boldsymbol{w}(\theta) = \sum_{\alpha_j \vDash \theta} \boldsymbol{w}(\alpha_j)$$

where $\vDash$ denotes the classical notion of logical implication. Conditional probabilities are defined in the usual manner

$$\boldsymbol{w}(\theta \mid \phi) = \frac{\boldsymbol{w}(\theta \wedge \phi)}{\boldsymbol{w}(\phi)}$$

when $\boldsymbol{w}(\phi) \neq 0$ and are left undefined otherwise.

[1] This condition is sometimes known as the Watts Assumption (see [28]).

We note that in this paper the use the term "belief function" will always denote a probability function in the above sense.

The first axiomatic framework referred to above stems from the notion of an *inference process* first formulated by Paris and Vencovská some twenty years ago (see [28], [27],[29], and [30]). The problematic of Paris and Vencovská is that of an isolated individual **A** whose belief function is in general not completely specified, but whose set of beliefs is instead regarded as a set of constraints **K** on the possible values which the vector $< w_1 \dots w_J >$ may take. The constraint set **K** therefore defines a certain region of the Euclidean space $\mathbb{R}^J$, denoted by $\mathbf{V_K}$, consisting of all vectors $< w_1 \dots w_J >$ which satisfy the constraints in **K** together with the conditions that $\sum_{j=1}^{J} w_j = 1$ and that $w_j \geq 0$ for all j. In the special case when **K** is the empty set of constraints, we denote the corresponding region $\mathbf{V_K}$ by $\mathbb{D}_J$.

It is assumed that the constraint sets **K** which we consider are consistent (i.e. $\mathbf{V_K}$ is non-empty), and are such that $\mathbf{V_K}$ has pleasant geometrical properties. More precisely, the exact requirement on a set of constraints **K** is that the set $\mathbf{V_K}$ forms a non-empty closed convex region of Euclidean space. Throughout the rest of this paper **all constraint sets to which we refer will be assumed to satisfy this requirement**, and we shall refer to such constraint sets as *nice* constraint sets[2] Paris and Vencovská ask the question: given any such **K**, by what rational principles should **A** choose his probabilistic belief function $\boldsymbol{w}$ consistent with **K** *in the absence of any other information*?

A set of constraints **K** as above is often called a *knowledge base*. A rule **I** which for every such **K** chooses such a $\boldsymbol{w} \in \mathbf{V_K}$ is called an *inference process*. Given **K** we denote the belief function $\boldsymbol{w}$ chosen by **I** by $\mathbf{I}(\mathbf{K})$. The question above can then be reformulated as: what self-evident general principles should an inference process **I** satisfy? This question has been intensively studied over the last twenty years and much is known. In particular in [27] Paris and Vencovská found an elegant set of principles which uniquely characterise the maximum entropy inference process[3], **ME**, which is defined as follows: given **K** as above, $\mathbf{ME}(\mathbf{K})$ chooses that unique belief function $\boldsymbol{w}$ which maximises the Shannon entropy of $\boldsymbol{w}$, defined as

$$-\sum_{j=1}^{J} w_j \log w_j$$

subject to the condition that $\boldsymbol{w} \in \mathbf{V_K}$. Although some of the principles used to characterise **ME** may individually be open to philosophical challenge, they are sufficiently convincing overall to give **ME** the appearance of a gold standard, in the sense that no other known inference process satisfies an equally convincing set of principles[4]. The Paris-Vencovská axiomatic characterisation of **ME** is particularly striking because it is quite independent of historically much earlier justifications of

[2]This formulation ensures that linear constraint conditions such as $\boldsymbol{w}(\theta) = a$, $\boldsymbol{w}(\phi \mid \psi) = b$, and $\boldsymbol{w}(\psi \mid \theta) \leq c$, where $a, b, c \in [0,1]$ and θ, ϕ, and ψ are Boolean combinations of the α_j's, are all permissible in **K** provided that the resulting constraint set **K** is consistent. Here a conditional constraint such as $\boldsymbol{w}(\psi \mid \theta) \leq c$ is interpreted as $\boldsymbol{w}(\psi \wedge \theta) \leq c\,\boldsymbol{w}(\theta)$ which is always a well-defined linear constraint, albeit vacuous when $\boldsymbol{w}(\theta) = 0$.. See e.g. [28] for further details.

[3]This characterisation considerably strengthens earlier work of [35].

[4]Other favored inference processes which satisfy many, but not all, of these principles are the minimum distance inference process, **MD**, the limit centre of mass process, $\mathbf{CM}^{\infty}$, all **Renyi** inference processes, and the remarkable **Maximin** process of [14]. (See Paris [28] for a general introduction to inference processes, and also [14], especially the comparative table in Chapter 9, for an excellent résumé of the current state of knowledge concerning this topic).

ME stemming either from interpretations of probability arising from ideas in statistical mechanics (see [15],[16], [26]), or from axiomatic treatments of the concept of information itself (as in [34], [31], [6]). While both of the latter kinds of treatment are conceptually attractive it may be argued that they carry more philosophical baggage than does a purely axiomatic treatment of the formal desiderata to be satisfied by an inference process itself .

We should note that in the framework of Paris, Vencovská, and their followers, the atoms $\alpha_1, \alpha_2, \ldots \alpha_J$ are usually taken to be the atoms of the Lindenbaum algebra for a finite language of the propositional calculus with propositional variables say $p_1 \ldots p_n$. In such a case J is 2^n and so is necessarily a power of 2. Such a presentation, which adds an extra semantic layer, has certain conceptual advantages in the formulation and justification of certain natural principles such as the language invariance and irrelevant information principles of [28]. However since we shall not consider principles of this type in the present paper , we shall assume that the mutually exclusive and exhaustive atoms $\alpha_1, \alpha_2, \ldots \alpha_J$ are given a priori, rather than being generated by the set of propositional variables of an underling language, and we may then allow J to take any positive integral value.

An apparently very different framework of probabilistic inference has been much studied in the decision theoretic literature. Given possible outcomes $\alpha_1, \alpha_2, \ldots \alpha_J$ as before, let $\{\mathbf{A}_i \mid i = 1 \ldots m\}$ be a finite set of agents each of whom possesses his own particular probabilistic belief function $\boldsymbol{w}^{(i)}$ on the set of outcomes, and let us suppose that these $\boldsymbol{w}^{(i)}$ have already been determined. How then should these individual belief functions be aggregated so as to yield a single probabilistic belief function $\boldsymbol{v}$ which most accurately represents the *collective* beliefs of the agents? We call such an aggregated belief function a *social belief function*, and a general method of aggregation a *pooling operator*. Again we can ask: what principles should a pooling operator satisfy? In this framework various plausible principles have been investigated extensively in the literature, and have in particular been used to characterise two popular, but very different pooling operators **LinOp** and **LogOp**. **LinOp** takes $\boldsymbol{v}$ to be the arithmetic mean of the $\boldsymbol{w}^{(i)}$, i.e.

$$v_j \;=\; \frac{1}{m}\sum_{i=1}^{m} w_j^{(i)} \quad \text{for each } \; j = 1 \ldots J$$

whereas **LogOp** chooses $\boldsymbol{v}$ to be the normalised geometric mean given by:

$$v_j \;=\; \frac{(\prod_{i=1}^{m} w_j^{(i)})^{\frac{1}{m}}}{\sum_{k=1}^{J}(\prod_{i=1}^{m} w_k^{(i)})^{\frac{1}{m}}} \quad \text{for each } \; j = 1 \ldots J$$

Various continua of other pooling operators related to **LinOp** and **LogOp** have also been investigated. However the existing axiomatic analysis of pooling operators, while technically simpler than the analysis of inference processes, is also more ambiguous and perhaps less intellectually satisfying in its conclusions than the analysis of inference processes developed within the Paris-Vencovská framework ; in the former case one arrives at rival, apparently plausible, axiomatic characterisations of various pooling operators, including in particular **LinOp** and **LogOp**, without any very convincing foundational criteria for deciding, within the limited context of the framework, which operator is justified, if any. (See [7],[12],[5],[8],[10],[11],[13],[25],[36],[37] for further discussion of the axiomatics of pooling operators).

In the present paper we seek to provide an axiomatic framework to extend the Paris-Vencovská notion of inference process to the multiagent case, thereby encompassing both the Paris-Vencovská framework of inference processes and the framework of pooling operators as special, or *marginal*, cases. To this end we consider, for any $m \geq 1$, a set $\mathbf{M}$ consisting of m individuals $\mathbf{A}_1 \ldots \mathbf{A}_m$, each of whom possesses his own nice set of constraints, respectively $\mathbf{K}_1 \ldots \mathbf{K}_m$, on his possible belief function on the set of outcomes $\{\alpha_1, \alpha_2, \ldots \alpha_J\}$. (Note that we are only assuming here that the beliefs of each individual are consistent, not that the beliefs of different individuals are jointly consistent). We shall refer to such a set $\mathbf{M}$ of individuals as a *college*. The intuitive problem now is how the college $\mathbf{M}$ should choose a single belief function which best represents the totality of information conveyed by $\mathbf{K}_1 \ldots \mathbf{K}_m$.

Definition 1.1

Let $\mathcal{C}$ denote a given fixed class of constraints sets. A *social inference process* for $\mathcal{C}$ is a function , $\mathfrak{F}$, which chooses, for any $m \geq 1$ and constraint sets $\mathbf{K}_1 \ldots \mathbf{K}_m \in \mathcal{C}$, a probability function on $\{\alpha_1, \alpha_2, \ldots \alpha_J\}$, denoted by $\mathfrak{F}(\mathbf{K}_1 \ldots \mathbf{K}_m)$, which we refer to as the *social belief function* defined by $\mathfrak{F}$ acting on $\mathbf{K}_1 \ldots \mathbf{K}_m$.

When considering general properties of unspecified social inference processes, we may not specify exactly what the class $\mathcal{C}$ is, but in general we shall always assume that $\mathcal{C}$ is a class of nice constraint sets.

Note that, trivially, provided that when $m = 1$ $\mathfrak{F}(\mathbf{K}) \in \mathbf{V_K}$ for all $\mathbf{K} \in \mathcal{C}$, $\mathfrak{F}$ marginalises to an inference process. On the other hand, in the special case where $\mathbf{K}_1 \ldots \mathbf{K}_m$ are such that $\mathbf{V}_{\mathbf{K}_i}$ is a singleton for all $i = 1 \ldots m$, then $\mathfrak{F}$ marginalises to a pooling operator. The new framework therefore encompasses naturally the two classical frameworks described above.

Again we can ask: what principles would we wish such a social inference process $\mathfrak{F}$ to satisfy *in the absence of any further information*? Is there any social inference process $\mathfrak{F}$ which satisfies them? If so, to which inference process and to which pooling operator does such an $\mathfrak{F}$ marginalise? It turns out that merely by posing these questions in the right framework, and by making certain simple mathematical observations, we can gain considerable insight. It is however essential to note that our standpoint is strictly that of a logician: we insist on the absoluteness of the qualification above that we are given no further information than that stated in the problem. In particular we are given no information about the expertise of the individuals or about the independence of their opinions. This insistence on sticking to a problem where the available information is rigidly defined is absolutely essential to our analysis, just as it is in the analysis of inference processes by Paris and Vencovská and their followers. We make no apology for the fact that such an assumption is almost always unrealistic: in order to tackle difficult foundational problems it is necessary to start with a general but *precisely defined* problematic. As has in essence been pointed out by Paris and Vencovská, unless one is prepared to make certain assumptions which precisely delimit the probabilistic information under consideration, even the classical notion of an inference process becomes incoherent. Indeed failure to define precisely the information framework lies behind several so-called paradoxes of reasoning under uncertainty[5].

[5]The interested reader may consult [30] for a detailed analysis of this point in connection with supposed paradoxes arising from "representation dependence".

2 An Axiomatic Framework for a Social Inference Process

The underlying idea of a social inference process is not new. (See e.g. [21],[25], [17],[18] for some specific ideas in the literature which are related, but formulated with far less generality). However, to the author's knowledge, the work which has been done hitherto has largely been pragmatically motivated, and has not considered foundational questions. This is possibly due in part to a rather tempting reductionism, which would see the problem of a finding a social inference process as a two stage process in which a classical inference process is first chosen and applied to the constraints $\mathbf{K}_i$ of each agent i to yield a belief function $\boldsymbol{w}^{(i)}$ appropriate to that agent, and a pooling operator is then chosen and applied to the set of $\boldsymbol{w}^{(i)}$ to yield a social belief function[6]. Of course from this reductionist point of view a social inference process would not be particularly interesting foundationally, since we could hardly expect an analysis of such social inference processes to tell us anything fundamentally new about collective probabilistic reasoning.

Our approach is in this respect completely different. We reject the two stage approach above on the grounds that the classical notion of an inference process applies to an isolated single individual, and is valid only on the assumption that that individual has absolutely no knowledge or beliefs other than those specified by his personal constraint set. Indeed the preliminary point should be made that in the case of an isolated individual $\mathbf{A}$, whereas $\mathbf{A}$'s constraint set $\mathbf{K}$ is subjective and personal to that individual, the *passage* from $\mathbf{K}$ to $\mathbf{A}$'s assumed belief function w via an inference process should be made using rational or normative principles, and should therefore be considered to have an *intersubjective* character. We should not confuse the epistemological status of w with that of $\mathbf{K}$. By hypothesis $\mathbf{K}$ represents the sum total of $\mathbf{A}$'s beliefs; *ipso facto* $\mathbf{K}$ also represents, in general, a description of the extent of $\mathbf{A}$'s *ignorance*. Thus while $\boldsymbol{w}$ may be regarded as the *belief function which best represents* $\mathbf{A}$'s subjective beliefs, it must not be confused with those beliefs themselves, since in the passage from $\mathbf{K}$ to $\boldsymbol{w}$ it is clear that certain "information" has been discarded[7]; thus, while $\boldsymbol{w}$ is determined by $\mathbf{K}$ once an inference process is given, neither $\mathbf{K}$ nor $\mathbf{V_K}$ can be recaptured from $\boldsymbol{w}$. As a trivial example we may note that specifying that $\mathbf{A}$'s constraint set $\mathbf{K}$ is empty, i.e. that $\mathbf{A}$ claims total ignorance, is informationally very different from specifying that $\mathbf{K}$ is such that $\mathbf{V_K} = \{< \frac{1}{J}, \frac{1}{J} \dots \frac{1}{J} >\}$, although the application of $\mathbf{ME}$, or of any other reasonable inference process, yields $w = < \frac{1}{J}, \frac{1}{J} \dots \frac{1}{J} >$ in both cases.

From this point of view the situation of an individual who is a member of a college whose members seek to *collaborate* together to elicit a social belief function seems quite different from that of an isolated individual. Indeed in the former context it appears more natural to assume as a normative principle that, if the social belief function is to be optimal, then each individual member $\mathbf{A}_i$ *should be deemed* to choose his personal belief function $\boldsymbol{w}^{(i)}$ so as to take account of the information provided by the other individuals, in such a way that $\boldsymbol{w}^{(i)}$ is *consistent* with his own belief set $\mathbf{K}_i$, while being as informationally close as possible to the social belief function $\mathfrak{F}(\mathbf{K}_1 \dots \mathbf{K}_m)$ which is to be defined. We will show in section 3

[6] We note that by no means are all authors reductionist in this sense: in particular although their concerns are somewhat different from ours, neither [25] nor [21] make such an assumption.

[7] The word "information" is used here in a different sense from that of Shannon information.

that this key idea is indeed mathematically coherent and can be used to define a particular social inference process with remarkable properties. Notice however that it is not necessary to assume that a given $\mathbf{A}_i$ subjectively or consciously holds the particular personal belief function $\boldsymbol{w}^{(i)}$ which is attributed to him by the procedure above: such an $\boldsymbol{w}^{(i)}$ is viewed as nothing more than the belief function which $\mathbf{A}_i$ *ought rationally to hold*, given the personal constraint set $\mathbf{K}_i$ which represents his own beliefs, together with the *extra information* available to him by virtue of his knowledge of the constraint sets of the remaining members of the college. Just as in the case of an isolated individual, the passage from $\mathbf{A}_i$'s actual subjective belief set $\mathbf{K}_i$ to his notional subjective belief function $\boldsymbol{w}^{(i)}$ has an intersubjective or normative character: however the calculation of $\boldsymbol{w}^{(i)}$ now depends not only on $\mathbf{K}_i$ but on the belief sets of all the other members of the college.

Considerations similar to the above give rise to an important general principle which a social inference process should satisfy, which we will call the collegial principle. However before we introduce this principle below, we shall first state some more obvious principles: mostly these are obvious transfers of familiar symmetry axioms from the theory of inference processes or from social choice theory.

The Equivalence Principle

If for all $i = 1 \ldots m$ $\mathbf{V}_{\mathbf{K}_i} = \mathbf{V}_{\mathbf{K'}_i}$ then

$$\mathfrak{F}(\mathbf{K}_1 \ldots \mathbf{K}_m) = \mathfrak{F}(\mathbf{K'}_1 \ldots \mathbf{K'}_m)$$

.

$\square$

Otherwise expressed the Equivalence Principle states that substituting constraint sets which are equivalent in the sense that the set of belief functions which satisfy them is unchanged will leave the values of $\mathfrak{F}$ invariant. This principle is a familiar one adopted from the theory of inference processes (cf. [28]). In this paper we shall always consider only social inference processes (or inference processes) which satisfy the Equivalence Principle. For this reason we may occasionally allow a certain sloppiness of notation in the sequel by identifying a constraint set $\mathbf{K}$ with its set of solutions $\mathbf{V}_{\mathbf{K}}$ where the meaning is clear and this avoids an awkward notation. In particular if $\boldsymbol{\Delta}$ is a non-empty closed convex set of belief functions then we may write $\mathbf{ME}(\boldsymbol{\Delta})$ to denote the unique $w \in \boldsymbol{\Delta}$ which maximises the Shannon entropy function.

The Anonymity Principle

This principle states that $\mathfrak{F}(\mathbf{K}_1 \ldots \mathbf{K}_m)$ depends only on the multiset of constraint sets $\{\mathbf{K}_1 \ldots \mathbf{K}_m\}$ and not on the characteristics of the individuals with which the $\mathbf{K}_i$'s are associated nor the order in which the $\mathbf{K}_i$'s are listed.

$\square$

The following natural principle ensures that $\mathfrak{F}$ does not choose a belief function which violates the beliefs of some member of the the college unless there is no alternative. The principle also ensures that $\mathfrak{F}$ behaves like a classical inference process in the special case when $m = 1$.

The Consistency Principle

If $\mathbf{K}_1 \ldots \mathbf{K}_m$ are such that
$$\bigcap_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i} \neq \emptyset$$
then
$$\mathfrak{F}(\mathbf{K}_1 \ldots \mathbf{K}_m) \in \bigcap_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}$$

□

The following principle is again a familiar one satisfied by classical inference processes (see [28]):

Let σ denote a permutation of the atoms of S. Such a σ induces a corresponding permutation on the coordinates of probability distributions $< w_1 \ldots w_J >$, and on the corresponding coordinates of variables occurring in the constraints of constraint sets $\mathbf{K}_i$, which we denote below with an obvious notation.

The Atomic Renaming Principle

For any permutation σ of the atoms of S, and for all $\mathbf{K}_1 \ldots \mathbf{K}_m$
$$\mathfrak{F}(\sigma(\mathbf{K}_1) \ldots \sigma(\mathbf{K}_m)) = \sigma(\mathfrak{F}(\mathbf{K}_1 \ldots \mathbf{K}_m))$$
.

□

We now state the important principle referred to earlier:

The Collegial Principle

A social inference process $\mathfrak{F}$ satisfies the Collegial Principle (abbreviated to *Collegiality*) if for any $m \geq 1$ and $\mathbf{A}_1 \ldots \mathbf{A}_m$ with respective constraint sets $\mathbf{K}_1 \ldots \mathbf{K}_m$, if for some $k < m$ $\mathfrak{F}(\mathbf{K}_1 \ldots \mathbf{K}_k)$ is consistent with $\mathbf{K}_{k+1} \cup \mathbf{K}_{k+2} \cup \ldots \cup \mathbf{K}_m$, then
$$\mathfrak{F}(\mathbf{K}_1 \ldots \mathbf{K}_m) = \mathfrak{F}(\mathbf{K}_1 \ldots \mathbf{K}_k)$$

□

Collegiality may be interpreted as stating the following: if the social belief function v generated by some subset of the college is consistent with the individual beliefs of the remaining members, then v is also the social belief function of the whole college. In particular this means that adding to the college a new individual whose constraint set is empty will leave the social belief function unchanged.

The consistency and collegiality principles together immediately imply that $\mathfrak{F}$ satisfies the following unanimity property:

Lemma 2.1

If $\mathfrak{F}$ satisfies Consistency and Collegiality then for any $\mathbf{K}$

$$\mathfrak{F}(\mathbf{K}\ldots\mathbf{K}) \;=\; \mathfrak{F}(\mathbf{K})\,.$$

$\square$

Our next axiom goes to the heart of certain basic intuitions concerning probability. For expository reasons we will consider first the case when $m = 1$, in which case we are essentially discussing a principle to be satisfied by a classical inference process. First we introduce some fairly obvious terminology.

Let $\boldsymbol{w}$ denote $\mathbf{A}_1$'s belief function. (Since we are considering the case when $m = 1$ we will drop the superscript from $\boldsymbol{w}^{(1)}$ for ease of notation). For some non-empty set of atoms $\{\alpha_{j_1}\ldots\alpha_{j_t}\}$ let ϕ denote the event $\bigvee_{r=1}^{t}\alpha_{j_r}$. Suppose that $\mathbf{K}$ denotes a set of constraints on the variables $w_{j_1}\ldots w_{j_t}$ which defines a non-empty closed convex region of t-dimensional Euclidean space with $\sum_{r=1}^{t} w_{j_r} \leq 1$ and all $w_{j_r} \geq 0$. We shall refer to such a $\mathbf{K}$ as a *nice set of constraints about* ϕ. Such a set of constraints $\mathbf{K}$ may also be thought of as a constraint set on the $\boldsymbol{w}$ which determines a closed convex region $\mathbf{V}_{\mathbf{K}}$ of $\mathbb{D}_J$ defined by

$$\mathbf{V}_{\mathbf{K}} \;=\; \{\boldsymbol{w} \mid < w_{j_1}\ldots w_{j_t} > \;\text{satisfies}\,\mathbf{K}\,\}.$$

Now let $\hat{w}_r$ denote $w(\alpha_{j_r} \mid \phi)$ for $r = 1\ldots t$, with the $\hat{w}_r$ undefined if $w(\phi) = 0$. Then $\hat{\boldsymbol{w}} = < \hat{w}_1\ldots\hat{w}_t >$ is a probability distribution provided that $w(\phi) \neq 0$. Let $\mathbf{K}$ be a nice set of constraints on the probability distribution $\hat{\boldsymbol{w}}$: we shall refer to such a $\mathbf{K}$ as a *nice set of constraints conditioned on* ϕ. In line with our previous conventions we shall consider such $\mathbf{K}$ to be trivially satisfied in the case when $\boldsymbol{w}(\phi) = 0$.

Again an important point here is that while a nice set of constraints $\mathbf{K}$ conditioned on ϕ as above is given as a set of constraints on $\hat{\boldsymbol{w}}$ it can equally well be interpreted as defining a certain equivalent set of constraints on $\boldsymbol{w}$ instead, and it is easy to see that, with a slight abuse of notation, the corresponding region $\mathbf{V}_{\mathbf{K}}$ of $\mathbb{D}_J$ defined by

$$\mathbf{V}_{\mathbf{K}} \;=\; \{\boldsymbol{w} \mid \hat{\boldsymbol{w}}\,\text{satisfies}\,\mathbf{K}\,\}$$

is both convex and closed.

In what follows we may regard both a nice set of constraints conditioned on some event ϕ, and a nice set of constraints about some event ϕ, *as if* they defined constraints on the probability function $\boldsymbol{w}$, as explained above.

Notice that while a nice set of constraints *conditioned* on ϕ can say nothing about the value of belief in ϕ itself, a nice set of constraints *about* ϕ may do so, and may even fix belief in ϕ at a particular value.

The following principle captures a basic intuition about probabilistic reasoning which is valid for all standard inference processes:

The Locality Principle (for an Inference Process)

An inference process $\mathbf{I}$ satisfies the locality principle if for all sentences ϕ and θ, every nice set of constraints $\mathbf{K}$ *conditioned on* ϕ, and every nice set of constraints $\mathbf{K}^*$ *about* $\neg\phi$,

$$\mathbf{I}(\mathbf{K} \cup \mathbf{K}^*)\,(\theta \mid \phi) \;=\; \mathbf{I}(\mathbf{K})\,(\theta \mid \phi)$$

provided that $\mathbf{I}(\mathbf{K} \cup \mathbf{K}^*)\,(\phi) \neq 0$ and $\mathbf{I}(\mathbf{K})\,(\phi) \neq 0$

□

Let us refer to the set of all events which logically imply the event ϕ as *the world of* ϕ. Then the Locality Principle may be roughly paraphrased as saying that if $\mathbf{K}$ contains only information about the *relative size* of probabilistic beliefs about events in the world of ϕ, while $\mathbf{K}^*$ contains only information about beliefs concerning events in the world of $\neg\phi$, then the values which the inference process $\mathbf{I}$ calculates for probabilities of events conditioned on ϕ should be unaffected by the information in $\mathbf{K}^*$, except in the trivial case when belief in ϕ is forced to take the value 0. Put rather more more succinctly: beliefs about the world of $\neg\phi$ should not affect beliefs conditioned on ϕ. Note that we cannot expect to satisfy a strengthened version of this principle which would have belief in the events in the world of ϕ unaffected by $\mathbf{K}^*$ since the constraints in $\mathbf{K}^*$ may well affect belief in ϕ itself. Thus the Locality Principle asserts that, *ceteris paribus,* rationally derived *relative probabilities between events inside a "world"* are unaffected by information about what happens strictly outside that world.

The Locality Principle is in essence a combination of both the Relativisation Principle[8] of Paris [28] and the Homogeneity Axiom of Hawes [14] . The following theorem, which demonstrates that the most commonly accepted inference processes all satisfy Locality, is very similar to results proved previously, especially to results in [14]. It follows from the theorem below that if we reject the Locality Principle for an inference process, then we are in effect forced to reject not just **ME**, but also all currently known plausible inference processes, including all inference processes derived by maximising a generalised notion of entropy. This is an important point heuristically when we come to extend the Locality Principle to the multiagent case.[9]

Theorem 2.2

The inferences processes **ME**, $\mathbf{CM}^{\infty}$, **MD** (minimum distance), together with all **Renyi**[10] inference processes, and the **Maximin** inference process of [14], all satisfy the Locality Principle.

[8]We note here that Csiszár in [3],[4], introduces a property which he calls locality, but which corresponds to the relativisation principle of Paris [28] and is much weaker than the notion of locality in the present paper.

[9]The proof of Theorem 2.2 is not however germane to understanding the remainder of this paper and may be safely be skipped if the reader so wishes.

[10]Renyi inference processes are those which maximise one of the family of generalised notions of entropy due to Alfred Renyi (see [31], [32], [14], [28], [23]). A definition of Renyi processes is given below in the proof of 2.2.

Proof:

Let $\boldsymbol{F}$ be a real valued function defined on the domain

$$\bigcup_{J \in \mathbb{N}^+} [0,1]^J$$

by

$$\boldsymbol{F}(\boldsymbol{w}) \; = \; \sum_{j=1}^{J} f(w_j) \qquad (1)$$

for some function $f : [0,1] \to \mathbb{R}$.

We will say that $\boldsymbol{F}$ is *deflation proof* if for every $J \in \mathbb{N}^+$, all $\boldsymbol{w}, \boldsymbol{u} \in \mathbb{D}_J$, and every $\lambda \in (0,1)$

$$\boldsymbol{F}(\lambda \boldsymbol{w}) \; < \; \boldsymbol{F}(\lambda \boldsymbol{v}) \;\text{ if and only if }\; \boldsymbol{F}(\boldsymbol{w}) \; < \; \boldsymbol{F}(\boldsymbol{v}) \qquad (2)$$

Here $\lambda \boldsymbol{w}$ denotes the scalar multiplication of $\boldsymbol{w}$ by λ . Note that $\lambda \boldsymbol{w}$ will not be a vector in $\mathbb{D}_J$ in the above case since its coordinates sum to λ instead of 1.

We will see below that any inference process $\mathbf{I}$ such that $\mathbf{I}(\mathbf{K})$ is defined to be that point $\boldsymbol{v} \in \mathbf{V_K}$ which maximises a strictly convex deflation proof function $\boldsymbol{F}$ of the above form satisfies the locality principle.

We first note the following lemma:

Lemma 2.3

The inference processes listed in the statement of Theorem 2.2, with the exception of $\mathbf{CM}^{\infty}$ and **Maximin** , may all be defined by the maximisation of deflation proof strictly convex functions of the form (1) above.

Proof:

The inference process **ME** is defined by maximising

$$\boldsymbol{F}(\boldsymbol{w}) \; = \; -\sum_{j=1}^{J} w_j \log w_j$$

subject to the given constraints. Now for $\boldsymbol{w} \in \mathbb{D}_J$

$$\boldsymbol{F}(\lambda \boldsymbol{w}) \; = \; -\sum_{j=1}^{J} \lambda w_j \log \lambda w_j \; = \; -\,\lambda \log \lambda \; + \; \lambda\, \boldsymbol{F}(\boldsymbol{w})$$

from which (2) follows at once.

The Renyi inference process $\mathbf{REN}_r$, where r is a fixed positive real parameter not equal to 1 , is given by maximising the function

$$\boldsymbol{F}(\boldsymbol{w}) \; = \; -\sum_{j=1}^{J} (w_j)^r$$

for $\boldsymbol{w} \in \mathbf{V_K}$ in the case when $r > 1$, and by maximising

$$\boldsymbol{F}(\boldsymbol{w}) \;=\; \sum_{j=1}^{J} (w_j)^r$$

for $\boldsymbol{w} \in \mathbf{V_K}$ in the case when $0 < r < 1$.

Since for the above functions $\boldsymbol{F}(\lambda \boldsymbol{w}) \;=\; \lambda^r \boldsymbol{F}(\boldsymbol{w})$, they also trivially satisfy (2) and so are deflation proof. Note that the minimum distance inference process **MD** is just $\mathbf{REN}_2$. The functions $\boldsymbol{F}$ defined above are all strictly convex (see e.g. [28])) and so the lemma follows.

Returning to the main proof, let **I** be an inference process such that $\mathbf{I}(\mathbf{K})$ is defined by the maximisation of a deflation proof strictly convex function $\boldsymbol{F}$ of the form as in (1) above. Let $\phi\ \theta$, $\mathbf{K}$, and $\mathbf{K}^*$ be as in the statement of the locality principle. Without loss of generality we may assume for notational convenience that the atoms are so ordered that for some k with $1 \leq k < J$

$$\phi \equiv \bigvee_{j=1}^{k} \alpha_j \quad \text{and} \quad \neg\phi \equiv \bigvee_{j=k+1}^{J} \alpha_j$$

Let $\boldsymbol{u} = \mathbf{SEP}(\mathbf{K})$ and let $\boldsymbol{v} = \mathbf{SEP}(\mathbf{K} \cup \mathbf{K}^*)$. Let $\boldsymbol{u}(\phi) \;=\; a$ and let $\boldsymbol{v}(\phi) \;=\; b$. By hypothesis we know that a and b are non-zero. It suffices for us to show that

$$< \frac{v_1}{b} \ldots \frac{v_k}{b} > \;=\; < \frac{u_1}{a} \ldots \frac{u_k}{a} > \tag{3}$$

Now notice that since the constraints of $\mathbf{K}^*$ refer only to coordinates $k+1 \ldots J$ while the constraints of $\mathbf{K}$ refer only to coordinates $1 \ldots k$, the solution $\boldsymbol{v}$ which by definition maximises $\sum_{j=1}^{J} f(w_j)$ subject to the condition that $\boldsymbol{w} \in \mathbf{V}_{\mathbf{K} \cup \mathbf{K}^*}$, must also satisfy the condition that $< v_1 \ldots v_k >$ is that vector $< w_1 \ldots w_k >$ which maximises $\sum_{j=1}^{k} f(w_j)$ subject to $< \frac{w_1}{b} \ldots \frac{w_k}{b} >$ satisfying the constraints of $\mathbf{K}$ together with the constraint that $\sum_{j=1}^{k} w_j \;=\; b$. Now changing variables by setting $y_j \;=\; \frac{w_j}{b}$, with $\boldsymbol{y} \;= < y_1 \ldots y_k >$ this is equivalent to maximising

$$\boldsymbol{F}(b\boldsymbol{y}) \;=\; \sum_{j=1}^{k} f(by_j)$$

subject to $\boldsymbol{y} \in \mathbb{D}_k$ and $\boldsymbol{y}$ satisfying the constraints of $\mathbf{K}$. However since $\boldsymbol{F}$ is deflation proof (and strictly convex) the unique $\boldsymbol{y} \in \mathbb{D}_k$ which achieves this maximisation does not depend on b and by setting $b \;=\; 1$ we see that it is just the unique vector $\boldsymbol{y} \in \mathbb{D}_k$ maximising $\boldsymbol{F}(\boldsymbol{y})$ and satisfying the constraints in $\mathbf{K}$. Since this definition is independent of both $\mathbf{K}^*$ and b, it follows by replacing $\mathbf{K}^*$ by the empty set of constraints and b by a that equation (3) holds, which completes the proof for the case of inference processes defined by the maximisation of a deflation proof strictly convex function of the form (1) above. By lemma 2.3 the theorem follows for all the inference processes mentioned except for $\mathbf{CM}^\infty$ and **Maximin**.

The fact that the limit centre of mass inference process, $\mathbf{CM}^\infty$, satisfies locality may either be proved using the standard definition of $\mathbf{CM}^\infty$ in [28], and slightly modifying the idea of the proof above, or simply by observing that by a result of Hawes [14], for any constraint set $\mathbf{K}$

$$\mathbf{CM}^\infty(\mathbf{K}) \;=\; lim_{r \to 0^+}\, \mathbf{REN}_r(\mathbf{K})$$

and then applying the results above already proved for $\mathbf{REN}_r$.
The result for **Maximin** also follows easily from results in Hawes [14]. This completes the proof of 2.2.

$\square$

While Theorem 2.2 above merely provides very strong corroborating evidence in favour of accepting the Locality Principle for an inference process, an interesting aspect of the *intuition* underlying the principle is that the justification for it appears no less cogent when we attempt to generalise it to the context of a social inference process. If we accept the intuition in favour of the Locality Principle in the case of a single individual then it is hard to see why we should reject analogous arguments in the case of a social belief function which is derived by considering the beliefs of m individuals each of whom has constraint sets of the type considered above. The argument is a general informational one: if information about probabilities conditioned on ϕ is unaffected by information about the world of $\neg\phi$, then, *ceteris paribus*, this should be true regardless of whether the information is obtained from one agent or from many agents. Accordingly we may formulate more generally

The Locality Principle (for a Social Inference Process)

For any $m \geq 1$ let $\mathbf{M}$ be a college of m individuals $\mathbf{A}_1 \ldots \mathbf{A}_m$. If for each $i = 1 \ldots m$ $\mathbf{K}_i$ is a nice set of constraints conditioned on ϕ, and $\mathbf{K}_i^*$ is a nice set of constraints about $\neg\phi$, then for every event θ

$$\mathfrak{F}(\mathbf{K}_1 \cup \mathbf{K}_1^*, \ldots, \mathbf{K}_m \cup \mathbf{K}_m^*)\,(\theta \mid \phi) \;=\; \mathfrak{F}(\mathbf{K}_1, \ldots, \mathbf{K}_m)\,(\theta \mid \phi)$$

provided that $\mathfrak{F}(\mathbf{K}_1 \cup \mathbf{K}_1^*, \ldots, \mathbf{K}_m \cup \mathbf{K}_m^*)\,(\phi) \neq 0$
and $\mathfrak{F}(\mathbf{K}_1, \ldots \mathbf{K}_m)\,(\phi) \neq 0$.

$\square$

At this point we make a simple observation. In the very special marginal case when for each i the constraint sets $\mathbf{K}_i \cup \mathbf{K}_i^*$ are such as to completely determine $\mathbf{A}_i$'s belief function, so that the task of $\mathfrak{F}$ reduces to that of a pooling operator, the locality principle above reduces to a condition closely related to the well-known condition on a pooling operator that it be *externally Bayesian.* [11] The Locality Principle may therefore be interpreted as a generalisation of a suitably formulated property of external Bayesianity. We will not discuss this further here except to note the important point that if $\mathfrak{F}$ is taken to satisfy Locality, then this fact alone seriously restricts those pooling operators to which it is possible for $\mathfrak{F}$ to marginalise. Thus while **LogOp** satisfies the relevant cases of Locality, as follows from Theorem 3.9 below, the popular pooling operator **LinOp** does not do so. To give a counterexample for **LinOp**, let $J = 3$, let $\theta = \alpha_1 \vee \alpha_2$ and let

$$\mathbf{K}_1 \;=\; \{\, w(\alpha_1 \mid \theta) = \tfrac{2}{3} \,\} \;\text{ and }\; \mathbf{K}_1^* \;=\; \{\, w(\neg\theta) = \tfrac{1}{4} \,\}$$
$$\mathbf{K}_2 \;=\; \{\, w(\alpha_1 \mid \theta) = \tfrac{1}{3} \,\} \;\text{ and }\; \mathbf{K}_2^* \;=\; \{\, w(\neg\theta) = \tfrac{5}{6} \,\}$$

Then the unique belief function satisfying $\mathbf{K}_1 \cup \mathbf{K}_1^*$ is $\boldsymbol{w}^{(1)} = < \frac{1}{2}, \frac{1}{4}, \frac{1}{4} >$ while the unique belief function satisfying $\mathbf{K}_2 \cup \mathbf{K}_2^*$ is $\boldsymbol{w}^{(2)} = < \frac{1}{18}, \frac{1}{9}, \frac{5}{6} >$.

[11] This condition was first formulated by Madansky [22] in 1964 and further analyzed in [9] and [13], where it is shown that the only externally Bayesian pooling operators are closely related to **LogOp**. See also [12] for related properties of pooling operators.

Applying **LinOp** we obtain

$$\mathbf{LinOp}(\mathbf{K}_1 \cup \mathbf{K}_1^*,\ \mathbf{K}_2 \cup \mathbf{K}_2^*)\,(\alpha_1 \mid \theta) \;=\; \tfrac{20}{33}$$

If we now set

$$\mathbf{K}_1^{**} \;=\; \{\, w(\neg\theta) = \tfrac{3}{4} \,\} \;\text{ and }\; \mathbf{K}_2^{**} \;=\; \{\, w(\neg\theta) = \tfrac{1}{2} \,\}$$

then the unique belief function satisfying $\mathbf{K}_1 \cup \mathbf{K}_1^{**}$ is $\boldsymbol{w}^{(1)} \;=<\tfrac{1}{6}, \tfrac{1}{12}, \tfrac{3}{4}>$ while the unique belief function satisfying $\mathbf{K}_2 \cup \mathbf{K}_2^{**}$ is $\boldsymbol{w}^{(2)} \;=<\tfrac{1}{6}, \tfrac{1}{3}, \tfrac{1}{2}>$.

Applying **LinOp** gives

$$\mathbf{LinOp}(\mathbf{K}_1 \cup \mathbf{K}_1^{**},\ \mathbf{K}_2 \cup \mathbf{K}_2^{**})\,(\alpha_1 \mid \theta) \;=\; \tfrac{4}{9} \;\neq\; \mathbf{LinOp}(\mathbf{K}_1 \cup \mathbf{K}_1^{*},\ \mathbf{K}_2 \cup \mathbf{K}_2^{*})\,(\alpha_1 \mid \theta)$$

showing that Locality fails for any $\mathfrak{F}$ which marginalises to the pooling operator **LinOp**.

By contrast it is easily verified that

$$\mathbf{LogOp}(\mathbf{K}_1 \cup \mathbf{K}_1^{**},\ \mathbf{K}_2 \cup \mathbf{K}_2^{**})\,(\alpha_1 \mid \theta) \;=\; \tfrac{1}{2} \;=\; \mathbf{LogOp}(\mathbf{K}_1 \cup \mathbf{K}_1^{*},\ \mathbf{K}_2 \cup \mathbf{K}_2^{*})\,(\alpha_1 \mid \theta)$$

as expected.

Related facts concerning **LinOp** and **LogOp** have been widely noted in the literature on pooling operators; what we are noting that is new here is that arguments in favour of the Locality Principle in the far broader context of a social inference process give a quite new perspective on the relative acceptability of classical pooling operators such as **LogOp** and **LinOp**.

Our final axiom relates to a hypothetical situation where several exact copies of a college are amalgamated into a single college.

A *clone* of a member $\mathbf{A}_i$ of $\mathbf{M}$ is a member $\mathbf{A}_{i'}$ whose set of belief constraints on his belief function is identical to that of $\mathbf{A}_i$: i.e. $\mathbf{K}_i = \mathbf{K}_{i'}$. Suppose now that each member $\mathbf{A}_i$ of $\mathbf{M}$ is replaced by n clones of $\mathbf{A}_i$, so that we obtain a new college $\mathbf{M^*}$ with nm members. $\mathbf{M^*}$ may equally be regarded as k copies of $\mathbf{M}$ amalgamated into a single college; so since the social belief function associated with each of these copies of $\mathbf{M}$ would be the same, we may argue that surely the result of amalgamating the copies into a single college $\mathbf{M^*}$ should again yield the same social belief function.

For any constraint set $\mathbf{K}$ let $n\mathbf{K}$ stand for a a sequence of n copies of $\mathbf{K}$. Then the heuristic argument above generates the following:

The Proportionality Principle

For any integer $n > 1$

$$\mathfrak{F}(n\mathbf{K}_1,\, n\mathbf{K}_2,\, \ldots\, ,n\mathbf{K}_m) \;=\; \mathfrak{F}(\mathbf{K}_1,\, \mathbf{K}_2,\, \ldots\, ,\mathbf{K}_m)$$

□

The proportionality principle looks rather innocent. Nevertheless we shall see in section 4 that a slight variant of the same idea formulated as a limiting version has some surprising consequences.

3 The Social Entropy Process (SEP)

In this section we introduce a natural social inference process, **SEP**, which extends both the inference process **ME** and the pooling operator **LogOp**. Our heuristic derivation of **SEP** will be purely information theoretic. We prove certain important structural properties necessary to show that **SEP** is well-defined, and we show in Theorem 3.9 that **SEP** satisfies the seven principles introduced in the previous section.

In order to avoid problems with our definition of **SEP** however, we are forced to add a slight further restriction to the set of m constraint sets $\mathbf{K}_1 \ldots \mathbf{K}_m$ which respectively represent the beliefs sets of the individuals $\mathbf{A}_1 \ldots \mathbf{A}_m$. We assume in this section that the constraints are such that there exists at least one atom α_{j_0} such that no constraint set $\mathbf{K}_i$ forces α_{j_0} to take belief 0. In the special case when each $\mathbf{K}_i$ specifies a unique probability distribution, the condition corresponds to that necessary to ensure that **LogOp** is well-defined.

In order to motivate the definition of **SEP** heuristically, let us imagine that the college $\mathbf{M}$ decide to appoint an independent chairman $\mathbf{A}_0$, whom we may suppose to be a mathematically trained philosopher, and whose only task is to aggregate the beliefs of $\mathbf{A}_1 \ldots \mathbf{A}_m$ into a social belief function $\boldsymbol{v}$ according to strictly rational criteria, but ignoring any personal beliefs which $\mathbf{A}_0$ himself may hold. He must then convince the members of $\mathbf{M}$ that his method is optimal.

$\mathbf{A}_0$ decides that as an initial criterion he will choose a social belief function $\boldsymbol{v} = \; < v_1 \ldots v_J >$ in such a manner as to minimize the average informational distance between $< v_1 \ldots v_J >$ and the m belief functions $\boldsymbol{w}^{(i)} = \; < w_1^{(i)} \ldots w_J^{(i)} >$ of the members of $\mathbf{M}$, where the $\boldsymbol{w}^{(i)}$ are each simultaneously chosen in such a manner as to minimize this quantity subject to the relevant sets of belief constraints $\mathbf{K}_i$ of each of the members of the college.

The standard measure of informational distance between probability distributions $\boldsymbol{v}$ and $\boldsymbol{u}$ is the well-studied notion of *cross-entropy*, sometimes known as Kullback-Leibler divergence, given by

$$\boldsymbol{CE}(\boldsymbol{v}, \boldsymbol{u}) = \sum_{j=1}^{J} v_j \log \frac{v_j}{u_j}$$

where the convention is observed that $v_j \log \frac{v_j}{u_j}$ takes the value 0 if $v_j = 0$ and the value $+\infty$ if $v_j \neq 0$ and $u_j = 0$.

We recall that cross-entropy is *not* a symmetric function; intuitively in the context of updating for a single agent $\boldsymbol{CE}(\boldsymbol{v},\boldsymbol{u})$ represents the informational distance from old belief function $\boldsymbol{u}$ to new belief function $\boldsymbol{v}$. Using this notion of informational distance $\mathbf{A}_0$'s idea is therefore to choose $\boldsymbol{v}$ and $\boldsymbol{w}^{(1)}\ldots\boldsymbol{w}^{(m)}$ with each $\boldsymbol{w}^{(i)}$ satisfying $\mathbf{K}_i$, so as to minimize

$$\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{CE}(\boldsymbol{v},\boldsymbol{w}^{(i)})$$

We will see below that while such a procedure will not by itself always produce unique belief functions for $\boldsymbol{v}$ and the associated $\boldsymbol{w}^{(1)}\ldots\boldsymbol{w}^{(m)}$, the set of possible belief functions satisfying these criteria has both a pleasant characterisation and a tight mathematical structure.

A fundamental property of cross-entropy which we shall need is

Lemma 3.1 The Gibbs Inequality

For all belief functions $\boldsymbol{v}$ and $\boldsymbol{u}$

$$\boldsymbol{CE}(\boldsymbol{v},\boldsymbol{u}) \geq 0$$

with equality holding if and only if $\boldsymbol{v}=\boldsymbol{u}$.

Proof: See [20] or [28].

□

The next lemma allows us to express $\mathbf{A}_0$'s criterion above in a much more convenient mathematical form.

Lemma 3.2

Let $\mathbf{K}_1\ldots\mathbf{K}_m$ be constraint sets on belief functions $\boldsymbol{w}^{(1)}\ldots\boldsymbol{w}^{(m)}$ respectively. Then the following are equivalent:

(i) The belief functions $\boldsymbol{v}$, $\boldsymbol{w}^{(1)},\ldots\boldsymbol{w}^{(m)}$ minimize the quantity

$$\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{CE}(\boldsymbol{v},\boldsymbol{w}^{(i)}) \tag{4}$$

subject to the given constraints.

(ii) The belief functions $\boldsymbol{w}^{(1)}\ldots\boldsymbol{w}^{(m)}$ maximize the quantity

$$\sum_{j=1}^{J}\left[\prod_{i=1}^{m}w_j^{(i)}\right]^{\frac{1}{m}} \tag{5}$$

subject to the given constraints, and

$$v_j = \frac{\left[\prod_{i=1}^{m}w_j^{(i)}\right]^{\frac{1}{m}}}{\sum_{j=1}^{J}\left[\prod_{i=1}^{m}w_j^{(i)}\right]^{\frac{1}{m}}} \tag{6}$$

for all $j=1\ldots J$.

Proof:

We note first that by our assumptions concerning the constraint sets, the minimum value of (4) must be finite. For by assumption there exists some j_0 and some $\boldsymbol{u}^{(i)} \in \mathbf{V}_{\mathbf{K}_i}$ such that $u_{j_0}^{(i)} \neq 0$ for all $i = 1 \ldots m$; then by replacing each $\boldsymbol{w}^{(i)}$ by $\boldsymbol{u}^{(i)}$ and setting $v_{j_0} = 1$ and all other v_j equal to zero gives (4) a finite value. From this it follows that for any j if v_j is non-zero then $w_j^{(i)}$ is non-zero for all $i = 1 \ldots m$. Thus we can rewrite (4) as

$$\sum_{j=1}^{J} v_j \log \frac{v_j}{\left[\prod_{i=1}^{m} w_j^{(i)}\right]^{\frac{1}{m}}}$$

or, equivalently, as

$$\sum_{j=1}^{J} v_j \log \frac{v_j}{\left(\frac{\left[\prod_{i=1}^{m} w_j^{(i)}\right]^{\frac{1}{m}}}{\sum_{j'=1}^{J}\left[\prod_{i=1}^{m} w_{j'}^{(i)}\right]^{\frac{1}{m}}}\right)} \quad - \quad \log \sum_{j'=1}^{J} \left[\prod_{i=1}^{m} w_{j'}^{(i)}\right]^{\frac{1}{m}} \tag{7}$$

which, by the Gibbs inequality, will for any given $\boldsymbol{w}^{(1)} \ldots \boldsymbol{w}^{(m)}$ take its minimum value when the first term vanishes and $\boldsymbol{v}$ is given by the expression at (6). On the other hand the second term is minimized when $\sum_{j'=1}^{J}\left[\prod_{i=1}^{m} w_{j'}^{(i)}\right]^{\frac{1}{m}}$ is maximized. It follows that the minimum possible value of (4) is obtained by *first* maximising $\sum_{j'=1}^{J}\left[\prod_{i=1}^{m} w_{j'}^{(i)}\right]^{\frac{1}{m}}$ subject to the constraints, and then letting $\boldsymbol{v}$ be determined by the equation (6).

□

The above lemma shows that Chairman $\mathbf{A}_0$'s initial criterion for selecting appropriate $\boldsymbol{v}$ for consideration as the social belief function can be reduced to the problem of finding those sequences of belief functions $\boldsymbol{w}^{(1)} \ldots \boldsymbol{w}^{(m)}$ which maximize $\sum_{j=1}^{J}\left[\prod_{i=1}^{m} w_j^{(i)}\right]^{\frac{1}{m}}$, subject to each $\boldsymbol{w}^{(i)}$ satisfying the relevant set of constraints $\mathbf{K}_i$. Notice that the function being maximized above is just a sum of geometric means. Since this function is bounded and continuous and the space over which it is being maximized is by assumption closed, a maximum value is certainly attained.

In order to make our presentation more readable we shall in future abbreviate $\mathbf{K}_1 \ldots \mathbf{K}_m$ by $\vec{\mathbf{K}}$.

Definition 3.3

For a sequence of constraint sets $\vec{\mathbf{K}}$ we define

$$M_{\vec{\mathbf{K}}} \;=\; Max\,\{\sum_{j=1}^{J}(\prod_{i=1}^{m} w_j^{(i)})^{\frac{1}{m}} \;\mid\; \boldsymbol{w}^{(i)} \in \mathbf{V}_{\mathbf{K}_i} \text{ for all } i = 1 \ldots m\,\}$$

□

It is now easy to see that

Lemma 3.4

Given constraint sets $\mathbf{K}_1 \ldots \mathbf{K}_m$ and $M_{\vec{\mathbf{K}}}$ defined as above then $0 < M_{\vec{\mathbf{K}}} \leq 1$. Furthermore the value $M_{\vec{\mathbf{K}}} = 1$ occurs if and only if for every $j = 1 \ldots J$ and for all $i, i' \in \{1 \ldots m\}$ $w_j^{(i)} = w_j^{(i')}$. Hence given $\mathbf{K}_1 \ldots \mathbf{K}_m$ the following are equivalent:

1. $M_{\vec{\mathbf{K}}} = 1$
2. Every $\boldsymbol{w}^{(1)} \ldots \boldsymbol{w}^{(m)}$ which generates the value $M_{\vec{\mathbf{K}}}$ satisfies $\boldsymbol{w}^{(1)} = \ldots = \boldsymbol{w}^{(m)} = \boldsymbol{v}$.
3. The constraint sets $\mathbf{K}_1 \ldots \mathbf{K}_m$ are jointly consistent: i.e there exists some belief function which satisfies all of them.

Proof:

Let $\boldsymbol{w}^{(1)} \ldots \boldsymbol{w}^{(m)}$ be belief functions satisfying $\mathbf{K}_1 \ldots \mathbf{K}_m$ respectively, and which generate the value $M_{\vec{\mathbf{K}}}$. First note that by assumption for some j_0 no $\mathbf{K}_i$ forces the probability given to atom α_{j_0} to be zero, and hence $M_{\vec{\mathbf{K}}} > 0$, since it is possible to choose belief functions $\boldsymbol{u}^{(1)} \ldots \boldsymbol{u}^{(m)}$ respectively consistent with $\mathbf{K}_1 \ldots \mathbf{K}_m$ such that $\left[\prod_{i=1}^m u_{j_0}^{(i)}\right]^{\frac{1}{m}} > 0$.
Now by applying the arithmetic-geometric mean inequality m times we get

$$M_{\vec{\mathbf{K}}} = \sum_{j=1}^{J} (\prod_{i=1}^{m} w_j^{(i)})^{\frac{1}{m}} \leq \sum_{j=1}^{J} \frac{1}{m} \sum_{i=1}^{m} w_j^{(i)} = 1 \quad \text{since} \quad \sum_{i=1}^{m} \sum_{j=1}^{J} w_j^{(i)} = m.$$

Moreover since equality for any of the arithmetic-geometric mean inequalities occurs just when all the terms are equal, the case $M_{\vec{\mathbf{K}}} = 1$ occurs if and only if $\boldsymbol{w}^{(1)} = \boldsymbol{w}^{(1)} = \ldots = \boldsymbol{w}^{(m)} = \boldsymbol{v}$. This suffices to prove the lemma.

□

Now it is obvious from the above that chairman $\mathbf{A}_0$'s proposed method of choosing $\boldsymbol{v}$ will not in general result in a uniquely defined social belief function. Indeed if $\bigcap_{i=1}^m \mathbf{V}_{\mathbf{K}_i} \neq \emptyset$ then *any* point $\boldsymbol{w}$ in this intersection, if adopted as the belief function of each member, will generate the maximum possible value for $M_{\vec{\mathbf{K}}}$ of 1 and so will be a possible candidate for a social belief function $\boldsymbol{v}$. Moreover even if $\bigcap_{i=1}^m \mathbf{V}_{\mathbf{K}_i} = \emptyset$ the process above may not result in a unique choice of either the $\boldsymbol{w}^{(i)}$ or of $\boldsymbol{v}$.

Chairman $\mathbf{A}_0$ now reasons as follows: if the result of the above operation of minimizing the average cross-entropy does not result in a unique solution for $\boldsymbol{v}$, then the best rational recourse which he has left is to choose that $\boldsymbol{v}$ which has maximum entropy from the set of possible $\boldsymbol{v}$ previously obtained, assuming of course that such a choice is well-defined. Chairman $\mathbf{A}_0$ reasons that by adopting this procedure he is treating the set of $\boldsymbol{v}$ defined by minimizing the average cross-entropy of $\boldsymbol{v}$ with possible belief functions of college members *as if* that were the set of his own possible belief functions, and then choosing a belief function from that set by applying the **ME** inference process, as he would if that were indeed the case.

However in order to show that this procedure is well-defined chairman $\mathbf{A}_0$ needs to prove certain technical results.

Definition 3.5

For constraint sets $\vec{\mathbf{K}}$ we define

$$\mathbf{\Gamma}(\vec{\mathbf{K}}) \;=\; \{< \boldsymbol{w}^{(1)} \ldots \boldsymbol{w}^{(m)} > \in \bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i} \;\mid\; \sum_{j=1}^{J} \left[\prod_{i=1}^{m} w_j^{(i)} \right]^{\frac{1}{m}} = M_{\vec{\mathbf{K}}} \,\}$$

By 3.2, each point $< \boldsymbol{w}^{(1)} \ldots \boldsymbol{w}^{(m)} >$ in $\mathbf{\Gamma}(\vec{\mathbf{K}})$ gives rise to a uniquely determined corresponding social belief function $\boldsymbol{v}$ whose j'th coordinate is given by

$$v_j \;=\; \frac{1}{M_{\vec{\mathbf{K}}}} (\prod_{i=1}^{m} w_j^{(i)})^{\frac{1}{m}}$$

We will refer to the $\boldsymbol{v}$ thus obtained from $< \boldsymbol{w}^{(1)} \ldots \boldsymbol{w}^{(m)} >$ as

$$\mathbf{LogOp}(\boldsymbol{w}^{(1)} \ldots \boldsymbol{w}^{(m)})$$

and we let

$$\mathbf{\Delta}(\vec{\mathbf{K}}) \;=\; \{\mathbf{LogOp}(\boldsymbol{w}^{(1)} \ldots \boldsymbol{w}^{(m)}) \;\mid < \boldsymbol{w}^{(1)} \ldots \boldsymbol{w}^{(m)} > \in \mathbf{\Gamma}(\vec{\mathbf{K}}) \,\}$$

□

$\mathbf{\Delta}(\vec{\mathbf{K}})$ is thus the candidate set of possible social belief functions from which Chairman $\mathbf{A}_0$ wishes to make his final choice by selecting the point in this set which has maximum entropy.

From now on we shall abbreviate a typical point $< \boldsymbol{w}^{(1)} \ldots \boldsymbol{w}^{(m)} >$ in $\bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}$ by $\vec{\boldsymbol{w}}$. For any such $\vec{\boldsymbol{w}}$ we denote the vector $< w_j^{(1)} \ldots w_j^{(m)} >$ by $\boldsymbol{w}_j$. Thus we may think of $\vec{\boldsymbol{w}}$ as an $m \times J$ matrix with rows $\boldsymbol{w}^{(i)}$, columns $\boldsymbol{w}_j$, and individual entries $w_j^{(i)}$.

Our problem is to analyze the linked structures of $\mathbf{\Gamma}(\vec{\mathbf{K}})$ and $\mathbf{\Delta}(\vec{\mathbf{K}})$, and in particular to show that $\mathbf{\Delta}(\vec{\mathbf{K}})$ is convex. A slight complicating factor in this analysis turns out to be the possibility that some entries in a matrix $\vec{\boldsymbol{w}} \in \mathbf{\Gamma}(\vec{\mathbf{K}})$ may turn out to be zero. Notice that the corresponding social belief function $\boldsymbol{v}$ will have j'th coordinate v_j equal to zero if and only if some entry in the column vector $\boldsymbol{w}_j$ is equal to zero. Such zero entries v_j may be classified as of two possible kinds: either $v_j = 0$ because for some i the constraint set $\mathbf{K}_i$ forces $w_j^{(i)} = 0$, or, when this is not the case, because for some i $w_j^{(i)} = 0$ just happens to be zero. The first case is in a certain sense trivial since for an arbitrary $\vec{\boldsymbol{w}} \in \bigotimes_{\boldsymbol{i=1}}^{\boldsymbol{m}} \mathbf{V}_{\mathbf{K}_{\boldsymbol{i}}}$ the columns $\boldsymbol{w}_j$ corresponding to such j will make zero contribution to the function to be maximised. For this reason it is convenient to introduce a notation which allows us to eliminate such j from consideration. Accordingly, for given $\vec{\mathbf{K}}$, we define the set of significant j, $\mathrm{Sig}_{\vec{\mathbf{K}}}$ by:

$$\mathrm{Sig}_{\vec{\mathbf{K}}} \;=\; \{\, j \mid \text{for no } i \text{ is it the case that } w_j^{(i)} \;=\; 0 \text{ for all } \boldsymbol{w}^{(i)} \in \mathbf{V}_{\mathbf{K}_i}\}$$

Notice that by our initial assumption about $\vec{\mathbf{K}}$ at the beginning of this section $\mathrm{Sig}_{\vec{\mathbf{K}}}$ is non-empty.

For any $\vec{\boldsymbol{w}} \in \bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}$ we now define $\vec{\boldsymbol{w}}_{\mathrm{Sig}_{\vec{\mathbf{K}}}}$ to be the projection of $\vec{\boldsymbol{w}}$ on to those coordinates (i,j) such that $j \in \mathrm{Sig}_{\vec{\mathbf{K}}}$; i.e. $\vec{\boldsymbol{w}}_{\mathrm{Sig}_{\vec{\mathbf{K}}}}$ may be viewed as the matrix obtained from the matrix $\vec{\boldsymbol{w}}$ by deleting those columns j for which $j \notin \mathrm{Sig}_{\vec{\mathbf{K}}}$. Similarly for any probability function $\boldsymbol{w}$ we define $\boldsymbol{w}_{\mathrm{Sig}_{\vec{\mathbf{K}}}}$ to be the projection of $\boldsymbol{w}$ to a vector obtained by deleting those coordinates which are not in $\mathrm{Sig}_{\vec{\mathbf{K}}}$. (Notice however that the effect of this is that the sum of the components of such a $\boldsymbol{w}_{\mathrm{Sig}_{\vec{\mathbf{K}}}}$ may be less than unity). Similarly we define

$$\boldsymbol{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}}) \;=\; \{\, \vec{\boldsymbol{w}}_{\mathrm{Sig}_{\vec{\mathbf{K}}}} \;\mid\; \vec{\boldsymbol{w}} \in \boldsymbol{\Gamma}(\vec{\mathbf{K}}) \,\}$$

and

$$\boldsymbol{\Delta}^{\mathrm{Sig}}(\vec{\mathbf{K}}) \;=\; \{\, \mathbf{v}_{\mathrm{Sig}_{\vec{\mathbf{K}}}} \;\mid\; \mathbf{v} \in \boldsymbol{\Delta}(\vec{\mathbf{K}}) \,\}$$

Note that in contrast to the situation for the row vectors of a matrix in $\boldsymbol{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$, the components of any vector in $\boldsymbol{\Delta}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ do sum to unity, and that there is therefore a trivial homeomorphism between $\boldsymbol{\Delta}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ and $\boldsymbol{\Delta}(\vec{\mathbf{K}})$.

The next theorem, which guarantees that Chairman $\mathbf{A}_0$'s plan is realisable, provides a crucial structure theorem for $\boldsymbol{\Gamma}(\vec{\mathbf{K}})$ and $\boldsymbol{\Delta}(\vec{\mathbf{K}})$, which depends strongly on the concavity properties of the geometric mean function and of sums of such functions.[12]

Theorem 3.6

Let $\vec{\mathbf{K}}$ be a fixed vector of constraint sets such that $\boldsymbol{\Delta}(\vec{\mathbf{K}})$ is not a singleton.

(i) Let $\vec{\boldsymbol{w}} \in \boldsymbol{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$, and let $\boldsymbol{v}$ be the corresponding point in $\boldsymbol{\Delta}^{\mathrm{Sig}}(\vec{\mathbf{K}})$.
Then for each $j \in \mathrm{Sig}_{\vec{\mathbf{K}}}$ then **either** $w_j^{(i)} = 0$ for all $i = 1 \ldots m$ **or** $w_j^{(i)}$ is nonzero for all $i = 1 \ldots m$.

Furthermore in the case when $w_j^{(i)}$ is nonzero for all $i = 1 \ldots m$, if $\overrightarrow{\boldsymbol{w'}}$ is any other point in $\boldsymbol{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ with corresponding point $\boldsymbol{v'}$ in $\boldsymbol{\Delta}^{\mathrm{Sig}}(\vec{\mathbf{K}})$, then

$$\boldsymbol{w}'_j \;=\; (1+\mu_j)\boldsymbol{w}_j \quad \text{for some } \mu_j \in \mathbb{R} \text{ with } \mu_j \geq -1.$$

and hence also

$$v'_j \;=\; (1+\mu_j)v_j$$

(ii) There is a point $\vec{\boldsymbol{w}} \in \boldsymbol{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ with corresponding $\boldsymbol{v} \in \boldsymbol{\Delta}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ such that for every other point $\overrightarrow{\boldsymbol{w'}} \in \boldsymbol{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ with corresponding $\boldsymbol{v'} \in \boldsymbol{\Delta}^{\mathrm{Sig}}(\vec{\mathbf{K}})$, for each $j \in \mathrm{Sig}_{\vec{\mathbf{K}}}$ there exists $\mu_j \geq -1$ such that

$$\boldsymbol{w}'_j \;=\; (1+\mu_j)\boldsymbol{w}_j$$

and

$$v'_j \;=\; (1+\mu_j)v_j$$

(iii) The regions $\boldsymbol{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$, $\boldsymbol{\Delta}^{\mathrm{Sig}}(\vec{\mathbf{K}})$, $\boldsymbol{\Gamma}(\vec{\mathbf{K}})$, and $\boldsymbol{\Delta}(\vec{\mathbf{K}})$ are all compact and convex.

[12] An earlier version of this theorem which was stated without proof in [38] contains an error because the statement of the result is incorrect for cases in which 0's appear in the coordinates.

(iv) If $\mathbf{LogOp}^{\text{Sig}}$ denotes the function defined on $\mathbf{\Gamma}^{\text{Sig}}(\vec{\mathbf{K}})$ by restricting the definition of the $\mathbf{LogOp}$ function defined on $\mathbf{\Gamma}(\vec{\mathbf{K}})$ in 3.5 above to those j which are in $\text{Sig}_{\vec{\mathbf{K}}}$, then

$$\mathbf{LogOp}^{\text{Sig}} : \ \mathbf{\Gamma}^{\text{Sig}}(\vec{\mathbf{K}}) \ \rightarrow \ \mathbf{\Delta}^{\text{Sig}}(\vec{\mathbf{K}})$$

is a continuous bijection.

Proof:

Define the function $\mathbf{F}: \ \bigotimes_{i=1}^{m} \mathbb{D}_J \ \rightarrow \ \mathbb{R}$: by

$$\mathbf{F}(\vec{\boldsymbol{w}}\,) \ = \ \sum_{j=1}^{J} \left[\prod_{i=1}^{m} w_j^{(i)} \right]^{\frac{1}{m}}$$

This is this function which is to be maximised for $\vec{\boldsymbol{w}} \ \in \ \bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}$ in order to define the points in the region $\mathbf{\Gamma}(\vec{\mathbf{K}})$. We note first of all that for non-negative arguments the geometric mean function is always concave (see e.g. [2]) and hence a sum of such functions is also concave. Since the region $\bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}$ is convex and compact by its definition, it follows that $\mathbf{F}$ attains a maximum value and hence that $\mathbf{\Gamma}(\vec{\mathbf{K}})$ is non-empty. Moreover it is an easy consequence of the definition of a concave function that the set of points which give maximal value to such a function over a compact convex region itself forms a compact convex set. Thus $\mathbf{\Gamma}(\vec{\mathbf{K}})$ is compact and convex. Since both compactness and convexity are preserved by projections in Euclidean space it follows that $\mathbf{\Gamma}^{\text{Sig}}(\vec{\mathbf{K}})$ is also compact and convex.

Let $[\bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}]^{\text{Sig}_{\vec{\mathbf{K}}}}$ denote the projection of $\bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}$ onto those coordinates with $j \ \in \ \text{Sig}_{\vec{\mathbf{K}}}$. This region is also compact and convex. Then if we define $\mathbf{F}^{\text{Sig}}$ for any $\vec{\boldsymbol{w}} \ \in \ \bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}$ by

$$\mathbf{F}^{\text{Sig}}(\vec{\boldsymbol{w}}_{\text{Sig}_{\vec{\mathbf{K}}}}\,) \ = \ \sum_{j \in \text{Sig}_{\vec{\mathbf{K}}}} (\prod_{i=1}^{m} w_j^{(i)})^{\frac{1}{m}}$$

then it is clear that

$$\mathbf{F}^{\text{Sig}}(\vec{\boldsymbol{w}}_{\text{Sig}_{\vec{\mathbf{K}}}}\,) \ = \ \mathbf{F}(\vec{\boldsymbol{w}}\,)$$

so that it suffices for us to confine our analysis to $\mathbf{F}^{\text{Sig}}$ acting on the points in $[\bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}]^{\text{Sig}_{\vec{\mathbf{K}}}}$.

Now let us consider a general point $\vec{\boldsymbol{a}} \ \in \mathbf{\Gamma}^{\text{Sig}}(\vec{\mathbf{K}})$. We will show that for every $j \ \in \ \text{Sig}_{\vec{\mathbf{K}}}$ we cannot have that $a_j^{(i)} \ = \ 0$ while $a_j^{(i')} \ \neq \ 0$ for some $i, i' \ \in \ \{1 \ldots m\}$. Suppose for contradiction that such j, i and i' exist. We first note that there exists some $\vec{\boldsymbol{b}} \ \in \ [\bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}]^{\text{Sig}_{\vec{\mathbf{K}}}}$ such that $b_j^{(i)} \ \neq \ 0$ for all $i \ = \ 1 \ldots m$ and all $j \ \in \ \text{Sig}_{\vec{\mathbf{K}}}$. This follows from the convexity of $[\bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}]^{\text{Sig}_{\vec{\mathbf{K}}}}$ since for each particular i and j we can by our assumptions choose some $\vec{\boldsymbol{x}} \ \in \ [\bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}]^{\text{Sig}_{\vec{\mathbf{K}}}}$ such that $x_j^{(i)} \neq 0$ and by convexity we can then form a suitable $\vec{\boldsymbol{b}}$ by taking the arithmetic mean of all these. So let us fix some such $\vec{\boldsymbol{b}}$.

Let $\vec{\boldsymbol{u}} \ = \ \vec{\boldsymbol{b}} - \vec{\boldsymbol{a}}$. Then by convexity, for any $\lambda \ \in \ [0,1]$, the point $\vec{\boldsymbol{a}} \ + \ \lambda\,\vec{\boldsymbol{u}}$ is in $[\bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}]^{\text{Sig}_{\vec{\mathbf{K}}}}$. Note that by the definition of $\vec{\boldsymbol{b}}$, for all i and j if $a_j^{(i)} = 0$

then $u_j^{(i)} > 0$.

Consider the behaviour of $\mathbf{F}^{\mathrm{Sig}}(\vec{\boldsymbol{a}} + \lambda\,\vec{\boldsymbol{u}})$ as $\lambda \to 0$. Now differentiating with respect to λ we get

$$\frac{d\mathbf{F}}{d\lambda}^{\mathrm{Sig}}(\vec{\boldsymbol{a}} + \lambda\,\vec{\boldsymbol{u}}) \;=\; \frac{1}{m}\sum_{j\in \mathrm{Sig}_{\vec{\mathbf{K}}}}\left[\left[\prod_{i=1}^{m}(a_j^{(i)} \,+\, \lambda u_j^{(i)})\right]^{\frac{1}{m}}\;\sum_{i=1}^{m}\frac{u_j^{(i)}}{a_j^{(i)} + \lambda u_j^{(i)}}\right]$$

As $\lambda \to 0^+$ we see that all terms on the right hand side are bounded except in the case of those i, j where $a_j^{(i)} = 0$ and at least one $a_j^{(i')}$ is non-zero for some $i' \neq i$, in which case that term tends to $+\infty$. Since we are supposing that such j, i and i' do exist, it follows that $\mathbf{F}^{\mathrm{Sig}}$ is increasing as $\vec{\boldsymbol{a}} + \lambda\,\vec{\boldsymbol{u}}$ moves away from $\vec{\boldsymbol{a}}$, and hence since $\mathbf{F}^{\mathrm{Sig}}$ is continuous at $\vec{\boldsymbol{a}}$, $\vec{\boldsymbol{a}}$ cannot be a maximum point of $[\bigotimes_{i=1}^{m}\mathbf{V}_{\mathbf{K}_i}]^{\mathrm{Sig}_{\vec{\mathbf{K}}}}$, contradicting hypothesis. Thus we have shown that for any point $\vec{\boldsymbol{w}}$ in $\mathbf{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ if some column vector of $\vec{\boldsymbol{w}}$ has a zero entry then that column vector is identically zero, which establishes the first part of (i).

The second part of (i) follows directly from (ii), so we will prove (ii) instead.
By (i) and the convexity of $\mathbf{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ there exists an $\vec{\boldsymbol{a}}$ such that if there exists *any* $\vec{\boldsymbol{b}}$ in $\mathbf{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ for which for some j in $\mathrm{Sig}_{\vec{\mathbf{K}}}$ $\boldsymbol{b}_j$ is not a zero vector then all the entries of $\boldsymbol{a}_j$ are non-zero. Let us fix such an $\vec{\boldsymbol{a}}$ and let $\vec{\boldsymbol{b}}$ be any other point in $\mathbf{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$. Again we consider $\vec{\boldsymbol{u}} = \vec{\boldsymbol{b}} - \vec{\boldsymbol{a}}$ for $\lambda \in [0,1]$, noting that in this case by the convexity of $\mathbf{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$, $\vec{\boldsymbol{a}} + \lambda\,\vec{\boldsymbol{u}}$ is a point of $\mathbf{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$, and hence $\mathbf{F}^{\mathrm{Sig}}\,(\vec{\boldsymbol{a}} + \lambda\,\vec{\boldsymbol{u}}) \;=\; \mathrm{M}_{\vec{\mathbf{K}}}$ and so has constant value.

Let $\mathrm{Sig}^*_{\vec{\mathbf{K}}}$ denote $\{j \mid j \in \mathrm{Sig}_{\vec{\mathbf{K}}} \text{ and } \boldsymbol{a}_j \neq \mathbf{0}\}$. Then by the definition of $\vec{\boldsymbol{a}}$ and of $\mathrm{Sig}^*_{\vec{\mathbf{K}}}$

$$\mathbf{F}^{\mathrm{Sig}}\,(\vec{\boldsymbol{a}} + \lambda\,\vec{\boldsymbol{u}}) \;=\; \sum_{j\in \mathrm{Sig}^*_{\vec{\mathbf{K}}}}(\prod_{i=1}^{m}(a_j^{(i)} \,+\, \lambda u_j^{(i)}))^{\frac{1}{m}}$$

Noting that all the $a_j^{(i)}$ occurring on the right are by definition non-zero, differentiating twice with respect to λ we have

$$\frac{d^2\mathbf{F}}{d\lambda^2}^{\mathrm{Sig}}(\vec{\boldsymbol{a}}+\lambda\vec{\boldsymbol{u}}) \;=\; \frac{1}{m^2}\sum_{j\in \mathrm{Sig}^*_{\vec{\mathbf{K}}}}\left[\prod_{i=1}^{m}(a_j^{(i)} \,+\, \lambda u_j^{(i)})\right]^{\frac{1}{m}}\left[\left[\sum_{i=1}^{m}\frac{u_j^{(i)}}{a_j^{(i)} + \lambda u_j^{(i)}}\right]^2 - \;m\sum_{i=1}^{m}\frac{(u_j^{(i)})^2}{(a_j^{(i)} + \lambda u_j^{(i)})^2}\right]$$

Since $\mathbf{F}^{\mathrm{Sig}}$ is constant for $\lambda \in [0,1]$, setting the above expression equal to 0 for $\lambda = 0$ we get

$$\frac{1}{m^2}\sum_{j\in \mathrm{Sig}^*_{\vec{\mathbf{K}}}}\left[\prod_{i=1}^{m}a_j^{(i)}\right]^{\frac{1}{m}}\left[\left[\sum_{i=1}^{m}\frac{u_j^{(i)}}{a_j^{(i)}}\right]^2 - \;m\sum_{i=1}^{m}\left[\frac{u_j^{(i)}}{a_j^{(i)}}\right]^2\right] \;=\; 0$$

from which we obtain

$$-\sum_{j\in \mathrm{Sig}^*_{\vec{\mathbf{K}}}}\left[\prod_{i=1}^{m}a_j^{(i)}\right]^{\frac{1}{m}}\sum_{i,i'=1\ldots m}\left[\frac{u_j^{(i)}}{a_j^{(i)}} \;-\; \frac{u_j^{(i')}}{a_j^{(i')}}\right]^2 \;=\; 0$$

From the negative definite form of the above expression we deduce that for all $j \in \mathrm{Sig}^*_{\vec{\mathbf{K}}}$ and all $i, i' = 1 \dots m$

$$\frac{u_j^{(i)}}{a_j^{(i)}} = \frac{u_j^{(i')}}{a_j^{(i')}}$$

whence for all $j \in \mathrm{Sig}^*_{\vec{\mathbf{K}}}$ and all $i, i' = 1 \dots m$

$$\frac{b_j^{(i)}}{a_j^{(i)}} = \frac{b_j^{(i')}}{a_j^{(i')}}$$

which suffices to establish 3.6(ii).

To show 3.6(iv) note that the function $\mathbf{LogOp}^{\mathrm{Sig}} : \mathbf{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}}) \to \mathbf{\Delta}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ is by definition continuous and surjective. However by (ii) it is also clearly injective. Finally to show 3.6(iii) we have already noted that $\mathbf{\Gamma}(\vec{\mathbf{K}})$ and $\mathbf{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ are compact and convex. Since $\mathbf{\Delta}(\vec{\mathbf{K}})$ and $\mathbf{\Delta}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ are the continuous images of these compact sets under $\mathbf{LogOp}$ and $\mathbf{LogOp}^{\mathrm{Sig}}$ respectively, it follows that $\mathbf{\Delta}(\vec{\mathbf{K}})$ and $\mathbf{\Delta}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ are also compact. From the convexity of $\mathbf{\Gamma}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ the convexity of $\mathbf{\Delta}^{\mathrm{Sig}}(\vec{\mathbf{K}})$ follows by (ii), while the convexity of $\mathbf{\Delta}(\vec{\mathbf{K}})$ follows immediately from that of $\mathbf{\Delta}^{\mathrm{Sig}}(\vec{\mathbf{K}})$. This completes the proof of 3.6.

□

Now since $\mathbf{\Delta}(\vec{\mathbf{K}})$ is a compact convex set by 3.6(iii) and since the entropy function

$$-\sum_{j=1}^{J} v_j \log v_j$$

is strictly concave and bounded over this set, the set contains a unique point $\boldsymbol{v}^{\mathbf{ME}}$ at which the entropy function achieves its maximum value. It follows at once that the following formal definition of the social inference process **SEP** defines, for every $\vec{\mathbf{K}}$ satisfying the conditions of this section, a unique social belief function.

Definition 3.7

The Social Entropy Process, **SEP**, is the social inference process defined by

$$\mathbf{SEP}(\vec{\mathbf{K}}) = \mathbf{ME}(\mathbf{\Delta}(\vec{\mathbf{K}}))$$

□

We remark that it follows immediately from the definition above that the social inference process **SEP** marginalises to the inference process **ME** and to the pooling operator **LogOp**.

It is worth noting that Theorem 3.6(i) at once provides a simple sufficient condition for $\mathbf{\Delta}(\vec{\mathbf{K}})$ to be a singleton and thus for the application of **ME** in the definition of $\mathbf{SEP}(\vec{\mathbf{K}})$ to be redundant:

Theorem 3.8

If $\mathbf{K}_1 \ldots \mathbf{K}_m$ are such that for each $j = 1 \ldots J$ except possibly at most one there exists some i with $1 \leq i \leq m$ such that the condition $\boldsymbol{w}^{(i)} \in \mathbf{V}_{\mathbf{K}_i}$ forces $w_j^{(i)}$ to take a unique value, then $\mathbf{\Delta}(\mathbf{K}_1 \ldots \mathbf{K}_m)$ is a singleton. In particular this occurs if for some i $\mathbf{V}_{\mathbf{K}_i}$ is a singleton.

□

Theorem 3.9

SEP satisfies the seven principles of the previous section: Equivalence, Anonymity, Atomic Renaming, Consistency, Collegiality, Locality, and Proportionality.

Proof:

The fact that principles of Equivalence, Anonymity, and Atomic Renaming hold for **SEP** follows easily from the basic symmetry properties of the definition of **SEP**.

To prove that **SEP** satifies Consistency, suppose that $\vec{\mathbf{K}} = \mathbf{K}_1 \ldots \mathbf{K}_m$ are such that

$$\bigcap_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i} \neq \emptyset$$

Then for any $\boldsymbol{u} \in \bigcap_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}$, if we set

$$\boldsymbol{v} = \boldsymbol{w}^{(1)} = \ldots = \boldsymbol{w}^{(m)} = \boldsymbol{u}$$

Then

$$\sum_{j=1}^{J} (\prod_{i=1}^{m} w_j^{(i)})^{\frac{1}{m}} = 1$$

and since by 3.4 $M_{\vec{\mathbf{K}}} \leq 1$, it follows that $M_{\vec{\mathbf{K}}} = 1$, and hence that $\boldsymbol{u} \in \mathbf{\Delta}(\vec{\mathbf{K}})$. Conversely by 3.4, since $M_{\vec{\mathbf{K}}} = 1$, then for any $\boldsymbol{v} \in \mathbf{\Delta}(\vec{\mathbf{K}})$ if some $\vec{\boldsymbol{w}} \in \mathbf{\Gamma}(\vec{\mathbf{K}})$ generates $\boldsymbol{v}$, then $\boldsymbol{v} = \boldsymbol{w}^{(1)} = \ldots = \boldsymbol{w}^{(m)}$, and so $\boldsymbol{v} \in \bigcap_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}$. It follows that

$$\mathbf{SEP}(\mathbf{K}_1 \ldots \mathbf{K}_m) \in \bigcap_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}$$

as required.

To prove Collegiality suppose that $\mathbf{K}_1 \ldots \mathbf{K}_m$ are such that for some k with $1 < k < m$

$$\mathbf{SEP}(\mathbf{K}_1 \ldots \mathbf{K}_k) \in \bigcap_{i=k+1}^{m} \mathbf{V}_{\mathbf{K}_i}$$

Let $\check{\boldsymbol{v}} = \mathbf{SEP}(\mathbf{K}_1 \ldots \mathbf{K}_k)$ and let $\hat{\boldsymbol{v}} = \mathbf{SEP}(\mathbf{K}_1 \ldots \mathbf{K}_m)$.
Let $< \check{\boldsymbol{w}}^{(1)} \ldots \check{\boldsymbol{w}}^{(k)} > \in \mathbf{\Gamma}(\mathbf{K}_1 \ldots \mathbf{K}_k)$ be such that $\check{\boldsymbol{v}} = \mathbf{LogOp}(\check{\boldsymbol{w}}^{(1)} \ldots \check{\boldsymbol{w}}^{(k)})$.

Similarly let $<\hat{\boldsymbol{w}}^{(1)}\ldots\hat{\boldsymbol{w}}^{(m)}>\in\boldsymbol{\Gamma}(\mathbf{K}_1\ \ldots\mathbf{K}_m)$ be such that $\hat{\boldsymbol{v}}\ =\ \mathbf{LogOp}(\hat{\boldsymbol{w}}^{(1)}\ldots\hat{\boldsymbol{w}}^{(m)})$ Then by definition

$$\sum_{i=1}^{k}\sum_{j=1}^{J}v_j\log\frac{v_j}{w_j^{(i)}}$$

takes its minimum possible value for $\boldsymbol{w}^{(1)}\ldots\boldsymbol{w}^{(k)}$ subject to the constraints $\mathbf{K}_1\ \ldots\mathbf{K}_k$ when $<\boldsymbol{w}^{(1)}\ldots\boldsymbol{w}^{(k)}>=<\breve{\boldsymbol{w}}^{(1)}\ldots\breve{\boldsymbol{w}}^{(k)}>$ and $\boldsymbol{v}\ =\ \mathbf{LogOp}(\breve{\boldsymbol{w}}^{(1)}\ldots\breve{\boldsymbol{w}}^{(k)})$. We denote this value by Min_1. Similarly

$$\sum_{i=1}^{m}\sum_{j=1}^{J}v_j\log\frac{v_j}{w_j^{(i)}}$$

takes its minimum possible value for $\boldsymbol{w}^{(1)}\ldots\boldsymbol{w}^{(m)}$ subject to the constraints $\mathbf{K}_1\ \ldots\mathbf{K}_m$ when $<\boldsymbol{w}^{(1)}\ldots\boldsymbol{w}^{(m)}>=<\hat{\boldsymbol{w}}^{(1)}\ldots\hat{\boldsymbol{w}}^{(m)}>$; and $\boldsymbol{v}\ =\ \mathbf{LogOp}(\hat{\boldsymbol{w}}^{(1)}\ldots\hat{\boldsymbol{w}}^{(m)})$. We denote this value by Min_2.

We now define $\breve{\boldsymbol{w}}^{(i)}$ to be equal to $\breve{\boldsymbol{v}}$ for $k+1\le i\le m$. Notice that by hypothesis $\breve{\boldsymbol{w}}^{(1)}\ldots\breve{\boldsymbol{w}}^{(m)}$ now satisfy respectively the constraints $\mathbf{K}_1\ \ldots\mathbf{K}_m$. Hence we have by the definitions above

$$Min_2\ \le\ \sum_{i=1}^{m}\sum_{j=1}^{J}\breve{v}_j\log\frac{\breve{v}_j}{\breve{w}_j^{(i)}}\ =\ \sum_{i=1}^{k}\sum_{j=1}^{J}\breve{v}_j\log\frac{\breve{v}_j}{\breve{w}_j^{(i)}}\ =\ Min_1. \tag{8}$$

Similarly we also have

$$Min_2\ =\ \sum_{i=1}^{m}\sum_{j=1}^{J}\hat{v}_j\log\frac{\hat{v}_j}{\hat{w}_j^{(i)}}\ \ge\ \sum_{i=1}^{k}\sum_{j=1}^{J}\hat{v}_j\log\frac{\hat{v}_j}{\hat{w}_j^{(i)}}\ \ge\ Min_1. \tag{9}$$

It follows that the six quantities appearing in (8) and (9) above are all equal, and hence that

$$\breve{\boldsymbol{v}}\ \text{and}\ \hat{\boldsymbol{v}}\ \text{are both in}\ \ \boldsymbol{\Delta}(\mathbf{K}_1\ \ldots\mathbf{K}_k)\cap\boldsymbol{\Delta}(\mathbf{K}_1\ \ldots\mathbf{K}_m).$$

However by definition $\breve{\boldsymbol{v}}$ is the unique belief function with the highest entropy in $\boldsymbol{\Delta}(\mathbf{K}_1\ \ldots\mathbf{K}_k)$, while $\hat{\boldsymbol{v}}$ is the unique belief function with the highest entropy in $\boldsymbol{\Delta}(\mathbf{K}_1\ \ldots\mathbf{K}_m)$. Hence $\breve{\boldsymbol{v}}\ =\ \hat{\boldsymbol{v}}$ as required.

To prove Locality, consider a college with members $\mathbf{A}_1\ldots\mathbf{A}_m$ initially having respective constraint sets $\mathbf{K}_1\ \ldots\mathbf{K}_m$, where each $\mathbf{K}_i$ is a nice set of constraints conditioned on some fixed non-contradictory sentence ϕ. Now for each $i=1\ldots m$ let $\mathbf{K}_i^*$ be a nice set of constraints about $\neg\phi$. We are given that $\mathbf{SEP}(\mathbf{K}_1\cup\mathbf{K}_1^*,\ \ldots,\ \mathbf{K}_m\cup\mathbf{K}_m^*)\,(\phi)\ \ne\ 0$ and that $\mathbf{SEP}(\mathbf{K}_1\,,\ldots\ \mathbf{K}_m)\,(\phi)\ \ne\ 0$. We must show that for any sentence θ

$$\mathbf{SEP}(\mathbf{K}_1\ \cup\ \mathbf{K}_1^*,\ \ldots\ ,\mathbf{K}_m\ \cup\ \mathbf{K}_m^*)\,(\theta\mid\phi)\ =\ \mathbf{SEP}(\mathbf{K}_1\,,\ldots\ ,\mathbf{K}_m)\,(\theta\mid\phi)\,.$$

Clearly for this purpose it suffice to show that for any atom α such that $\alpha\models\phi$

$$\mathbf{SEP}(\mathbf{K}_1\ \cup\ \mathbf{K}_1^*,\ \ldots\ ,\mathbf{K}_m\ \cup\ \mathbf{K}_m^*)\,(\alpha\mid\phi)\ =\ \mathbf{SEP}(\mathbf{K}_1\,,\ldots\ ,\mathbf{K}_m)\,(\alpha\mid\phi)\,.$$

Notice that while we assume about each $\mathbf{K}_i$ that it determines a closed convex set of probability functions *conditioned* on ϕ, such a $\mathbf{K}_i$ when interpreted as a set of constraints about beliefs in the original atoms $\alpha_1,\alpha_2,\ldots\alpha_J$ also determines a closed convex region of $\mathbb{D}_J$ which as usual we denote by $\mathbf{V}_{\mathbf{K}_i}$. Hence $\mathbf{V}_{\mathbf{K}_i\cup\mathbf{K}_i^*}$ is also a closed convex region of $\mathbb{D}_J$. Furthermore the conditions imply that for each $i = 1 \ldots m$ $\mathbf{K}_i \cup \mathbf{K}_i^*$, is consistent, and hence the above applications of **SEP** are legitimately made.

Without loss of generality we may assume as in the proof of 2.2 that the atoms are so ordered that for some k with $1 \leq k < J$

$$\phi \equiv \bigvee_{j=1}^{k} \alpha_j \quad \text{and} \quad \neg\phi \equiv \bigvee_{j=k+1}^{J} \alpha_j$$

Let $\boldsymbol{u} = \mathbf{SEP}(\mathbf{K}_1, \ldots \mathbf{K}_m)$ be generated by $\vec{\boldsymbol{x}} \in \mathbf{\Gamma}(\mathbf{K}_1, \ldots, \mathbf{K}_m)$, and let $\boldsymbol{v} = \mathbf{SEP}(\mathbf{K}_1 \cup \mathbf{K}_1^*, \ldots, \mathbf{K}_m \cup \mathbf{K}_m^*)$ be generated by $\vec{\boldsymbol{y}} \in \mathbf{\Gamma}(\mathbf{K}_1 \cup \mathbf{K}_1^*, \ldots, \mathbf{K}_m \cup \mathbf{K}_m^*)$. For each $i = 1 \ldots m$, let $\sum_{j=1}^{k} x_j^{(i)} = a^{(i)}$, and let $\sum_{j=1}^{k} y_j^{(i)} = b^{(i)}$. Note that $a^{(i)}$ and $b^{(i)}$ are non-zero for all i since otherwise ϕ would get social belief zero contradicting hypotheses.

Now consider the point $\vec{\boldsymbol{z}} \in \bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}$ given for each $i = 1 \ldots m$ by

$$z_j^{(i)} = \begin{cases} y_j^{(i)} \frac{a^{(i)}}{b^{(i)}} & \text{for } j = 1, \ldots, k \\ x_j^{(i)} & \text{for } j = k+1, \ldots, J \end{cases}$$

By the definition of the point $\vec{\boldsymbol{x}}$ we know that

$$\sum_{j=1}^{J} \left[\prod_{i=1}^{m} z_j^{(i)} \right]^{\frac{1}{m}} \leq \sum_{j=1}^{J} \left[\prod_{i=1}^{m} x_j^{(i)} \right]^{\frac{1}{m}}$$

from which it follows that

$$\sum_{j=1}^{k} \left[\prod_{i=1}^{m} y_j^{(i)} \frac{a^{(i)}}{b^{(i)}} \right]^{\frac{1}{m}} \leq \sum_{j=1}^{k} \left[\prod_{i=1}^{m} x_j^{(i)} \right]^{\frac{1}{m}}$$

Dividing both sides by $\left[\prod_{i=1}^{m} a^{(i)} \right]^{\frac{1}{m}}$ we obtain that

$$\sum_{j=1}^{k} \left[\prod_{i=1}^{m} \frac{y_j^{(i)}}{b^{(i)}} \right]^{\frac{1}{m}} \leq \sum_{j=1}^{k} \left[\prod_{i=1}^{m} \frac{x_j^{(i)}}{a^{(i)}} \right]^{\frac{1}{m}} .$$

However by repeating a similar argument, but this time with $\vec{\boldsymbol{x}}$ and $\vec{\boldsymbol{y}}$ interchanged we obtain the reverse inequality, from which it follows that

$$\sum_{j=1}^{k} \left[\prod_{i=1}^{m} \frac{y_j^{(i)}}{b^{(i)}} \right]^{\frac{1}{m}} = \sum_{j=1}^{k} \left[\prod_{i=1}^{m} \frac{x_j^{(i)}}{a^{(i)}} \right]^{\frac{1}{m}} = M_1 \text{ say.} \tag{10}$$

Note that the above equality implies that the value M_1 does not depend on the $\mathbf{K}_i^*$ in any way.

Let $\sum_{j=k+1}^{J} \left[\prod_{i=1}^{m} y_j^{(i)} \right]^{\frac{1}{m}} = M_2$ and let $\left[\prod_{i=1}^{m} b^{(i)} \right]^{\frac{1}{m}} = B$.

Then from (3) we know that $\sum_{j=1}^{k} \left[\prod_{i=1}^{m} y_j^{(i)} \right]^{\frac{1}{m}} = M_1 B$.

Let us denote by C the quantity

$$\sum_{j=1}^{J}\left[\prod_{i=1}^{m} y_j^{(i)}\right]^{\frac{1}{m}} \;=\; M_1 B \;+\; M_2 \tag{11}$$

and we note that by definition C is the maximal value which can be taken by $\sum_{j=1}^{J}\left[\prod_{i=1}^{m} t_j^{(i)}\right]^{\frac{1}{m}}$ for any $\vec{\boldsymbol{t}} \in \bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i \cup \mathbf{K}_i^*}$. We now consider those $\vec{\boldsymbol{t}}$ of this form for which $t_j^{(i)} = y_j^{(i)}$ for all $j = k+1, \ldots, J$ and all $i = 1, \ldots, m$. Then, since for each $j = 1 \ldots k$ $\quad v_j = C^{-1}\left[\prod_{i=1}^{m} y_j^{(i)}\right]^{\frac{1}{m}}$ the definition of $\vec{\boldsymbol{y}}$ ensures that the column vectors $\boldsymbol{y}_1 \ldots \boldsymbol{y}_k$ are of the form $\boldsymbol{t}_1 \ldots \boldsymbol{t}_k$ where

$$-\sum_{j=1}^{k} C^{-1}\left[\prod_{i=1}^{m} t_j^{(i)}\right]^{\frac{1}{m}} \log\left(C^{-1}\left[\prod_{i=1}^{m} t_j^{(i)}\right]^{\frac{1}{m}}\right) \tag{12}$$

is maximised subject to the conditions that for each i the probability distribution $< \frac{t_1^{(i)}}{b^{(i)}} \ldots \frac{t_k^{(i)}}{b^{(i)}} >$ satisfies the constraint set $\mathbf{K}_i$, that

$$\sum_{j=1}^{k}\left[\prod_{i=1}^{m} t_j^{(i)}\right]^{\frac{1}{m}} \;=\; M_1 B \tag{13}$$

and that for each i

$$\sum_{j=1}^{k} t_j^{(i)} \;=\; b^{(i)} \quad . \tag{14}$$

Using some elementary algebra and (13) above we can rewrite the quantity in (12) which is to be maximised as

$$-\,\frac{M_1 B}{C}\,\log\frac{B}{C} \;-\; \frac{B}{C}\sum_{j=1}^{k}\left[\prod_{i=1}^{m}\frac{t_j^{(i)}}{b^{(i)}}\right]^{\frac{1}{m}}\log\left[\prod_{i=1}^{m}\frac{t_j^{(i)}}{b^{(i)}}\right]^{\frac{1}{m}} \tag{15}$$

Now since B , C , and M_1 , are positive constants for the $\vec{\boldsymbol{t}}$ under consideration, it follows that maximising (15), or equivalently (12), under the given constraints, is equivalent to maximising

$$-\;\sum_{j=1}^{k}\left[\prod_{i=1}^{m}\frac{t_j^{(i)}}{b^{(i)}}\right]^{\frac{1}{m}}\log\left[\prod_{i=1}^{m}\frac{t_j^{(i)}}{b^{(i)}}\right]^{\frac{1}{m}} \tag{16}$$

Hence, writing $w_j^{(i)}$ for $\frac{t_j^{(i)}}{b^{(i)}}$, if follows that this is in turn equivalent to maximising

$$-\;\sum_{j=1}^{k}\left[\prod_{i=1}^{m} w_j^{(i)}\right]^{\frac{1}{m}}\log\left[\prod_{i=1}^{m} w_j^{(i)}\right]^{\frac{1}{m}} \tag{17}$$

subject to the constraints that each k-dimensional row vector $\boldsymbol{w}^{(i)} \;=< w_1^{(i)} \ldots w_k^{(i)} >$ sums to 1 and satisfies $\mathbf{K}_i$ when interpreted as a probability function conditioned on ϕ , and that

$$\sum_{j=1}^{k}\left[\prod_{i=1}^{m} w_j^{(i)}\right]^{\frac{1}{m}} \;=\; M_1 \tag{18}$$

Now by the remark following (10), the value M_1 must be the largest possible which can be attained by $\sum_{j=1}^{k}\left[\prod_{i=1}^{m} w_j^{(i)}\right]^{\frac{1}{m}}$ for the $\boldsymbol{w}^{(i)}$ probability functions satisfying the $\mathbf{K}_i$. Hence since the $\mathbf{K}_i$ are nice constraint sets, it follows by the fact that **SEP** is well-defined that any solution for $\vec{\boldsymbol{w}}$ to the above maximisation problem generates the unique $\mathbf{SEP}(\mathbf{K}_1, \ldots, \mathbf{K}_m)$ solution given by

$$\mathbf{SEP}(\mathbf{K}_1, \ldots, \mathbf{K}_m)(\alpha_j \mid \phi) = \frac{\left[\prod_{i=1}^{m} w_j^{(i)}\right]^{\frac{1}{m}}}{\sum_{r=1}^{k}\left[\prod_{i=1}^{m} w_r^{(i)}\right]^{\frac{1}{m}}}$$

for $j = 1 \ldots k$.
However by the definition of the above $w_j^{(i)}$ and the uniqueness of the **SEP** values, it follows that for such a solution $\vec{\boldsymbol{w}}$, for each $j = 1 \ldots k$

$$\left[\prod_{i=1}^{m} w_j^{(i)}\right]^{\frac{1}{m}} = \left[\prod_{i=1}^{m} \frac{y_j^{(i)}}{b^{(i)}}\right]^{\frac{1}{m}} \tag{19}$$

whence for each $j = 1 \ldots k$

$$\begin{aligned}
\mathbf{SEP}(\mathbf{K}_1, \ldots, \mathbf{K}_m)(\alpha_j \mid \phi) &= \frac{\left[\prod_{i=1}^{m} \frac{y_j^{(i)}}{b^{(i)}}\right]^{\frac{1}{m}}}{\sum_{r=1}^{k}\left[\prod_{i=1}^{m} \frac{y_r^{(i)}}{b^{(i)}}\right]^{\frac{1}{m}}} \\
&= \frac{C^{-1}\left[\prod_{i=1}^{m} y_j^{(i)}\right]^{\frac{1}{m}}}{C^{-1}\sum_{r=1}^{k}\left[\prod_{i=1}^{m} y_r^{(i)}\right]^{\frac{1}{m}}} \\
&= \mathbf{SEP}(\mathbf{K}_1 \cup \mathbf{K}_1^*, \ldots, \mathbf{K}_m \cup \mathbf{K}_m^*)(\alpha_j \mid \phi)
\end{aligned}$$

as required. This concludes the proof of Locality.

It remains for us to prove that **SEP** satisfies Proportionality.

Let $\mathbf{K}_1, \ldots, \mathbf{K}_m$ be constraint sets and for each $r = 1 \ldots n$ let $\mathbf{K}_{ir}$ denote a copy of the constraint set $\mathbf{K}_i$, so that $\mathbf{V}_{\mathbf{K}_{ir}} = \mathbf{V}_{\mathbf{K}_i}$. As a shorthand we denote the sequence $\mathbf{K}_{i1} \ldots \mathbf{K}_{in}$ by $n\mathbf{K}_i$. Clearly it suffices for us to prove that

$$\boldsymbol{\Delta}(n\mathbf{K}_1, n\mathbf{K}_2, \ldots, n\mathbf{K}_m) = \boldsymbol{\Delta}(\mathbf{K}_1, \mathbf{K}_2, \ldots, \mathbf{K}_m) \tag{20}$$

Let $\boldsymbol{v} \in \boldsymbol{\Delta}(n\mathbf{K}_1, n\mathbf{K}_2, \ldots, n\mathbf{K}_m)$ be generated by some $\vec{\boldsymbol{w}} \in \boldsymbol{\Gamma}(n\mathbf{K}_1, n\mathbf{K}_2, \ldots, n\mathbf{K}_m)$. Then letting

$$\sum_{r=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{J} v_j \log\frac{v_j}{w_j^{(ir)}} = D \tag{21}$$

by definition D is minimal subject only to the constraints that $\boldsymbol{w}^{(ir)} \in \mathbf{V}_{\mathbf{K}_i}$ for all $r = 1 \ldots n$ and $i = 1 \ldots m$, (but no constraints on $\boldsymbol{v}$). Then for each $r = 1 \ldots n$

$$\sum_{i=1}^{m}\sum_{j=1}^{J} v_j \log\frac{v_j}{w_j^{(ir)}} = \frac{D}{n} \tag{22}$$

and

$\frac{D}{n}$ is the minimum value which can be taken by

$$\sum_{i=1}^{m} \sum_{j=1}^{J} v_j \log \frac{v_j}{z_j^{(i)}} \quad \text{for } \boldsymbol{z}^{(i)} \in \mathbf{V}_{\mathbf{K}_i} \tag{23}$$

(22) holds because otherwise we would have that for some r_0 with $1 \leq r_0 \leq n$

$$\sum_{i=1}^{m} \sum_{j=1}^{J} v_j \log \frac{v_j}{w_j^{(ir_0)}} < \frac{D}{n}$$

and if we then define $\vec{\boldsymbol{y}}$ by

$$y_j^{(ir)} = w_j^{(ir_0)}$$

for all i, j and r, we would have that $\sum_{r=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{J} v_j \log \frac{v_j}{y_j^{(ir)}} < D$ contradicting the definition of D in (21). The same argument shows also that (23) holds. From (22) and (23) it follows that $\boldsymbol{v} \in \boldsymbol{\Delta}(\mathbf{K}_1, \mathbf{K}_2, \ldots, \mathbf{K}_m)$.

Conversely if some $\boldsymbol{v} \in \boldsymbol{\Delta}(\mathbf{K}_1, \mathbf{K}_2, \ldots, \mathbf{K}_m)$ is generated by a $\vec{\boldsymbol{z}} \in \boldsymbol{\Gamma}(\mathbf{K}_1, \mathbf{K}_2, \ldots, \mathbf{K}_m)$ then it is easy to see that

$$\sum_{i=1}^{m} \sum_{j=1}^{J} v_j \log \frac{v_j}{z_j^{(i)}} = \frac{D}{n} \tag{24}$$

where D is the minimal value defined at (14) since the value of $\sum_{i=1}^{m} \sum_{j=1}^{J} v_j \log \frac{v_j}{z_j^{(i)}}$ cannot be smaller than $\frac{D}{n}$ by the same argument used to show (22) and (23). However if we now define $\vec{\boldsymbol{w}}$ by $w_j^{(ir)} = z_j^{(i)}$ then by (24) the equation (21) holds and so $\boldsymbol{v} \in \boldsymbol{\Delta}(n\mathbf{K}_1, n\mathbf{K}_2, \ldots, n\mathbf{K}_m)$.
Thus $\boldsymbol{\Delta}(n\mathbf{K}_1, n\mathbf{K}_2, \ldots, n\mathbf{K}_m) = \boldsymbol{\Delta}(\mathbf{K}_1, \mathbf{K}_2, \ldots, \mathbf{K}_m)$ as required.

This concludes the proof of Theorem 3.9.

□

We remark that Sam Savage [33] has shown that a certain form of converse of Collegiality holds for **SEP**. Namely, if $\mathbf{K}_1 \ldots \mathbf{K}_m$ are such that for each $j = 1 \ldots J$ $\mathbf{SEP}(\mathbf{K}_1 \ldots \mathbf{K}_m)(\alpha_j) \neq 0$, then if $\mathbf{SEP}(\mathbf{K}_1 \ldots \mathbf{K}_{m-1}) \notin \mathbf{V}_{\mathbf{K}_m}$ then $\mathbf{SEP}(\mathbf{K}_1 \ldots \mathbf{K}_{m-1}) = \mathbf{SEP}(\mathbf{K}_1 \ldots \mathbf{K}_m)$.

□

We end this section with some brief remarks concerning possible generalisations to the context of a social inference process, and in particular to **SEP**, of the remaining key principles which were identified by Paris and Vencovská ([28],[27]) as characterising the **ME** inference process.

One such key principle satisfied by **ME**, is that of **Open-Mindedness**. An inference process **I** satisfies Open-Mindedness if for every constraint set **K**, for all $j = 1 \ldots J$ $\mathbf{I}(\mathbf{K})(\alpha_j) \neq 0$ unless $w_j = 0$ for all $\boldsymbol{w} \in \mathbf{V}_{\mathbf{K}}$. The most obvious way of extending this principle to the case of a social inference process $\mathfrak{F}$ would seem to be to propose that for all $j = 1 \ldots J$ and for all $\mathbf{K}_1, \mathbf{K}_2, \ldots \mathbf{K}_m$ $\mathfrak{F}(\mathbf{K}_1, \mathbf{K}_2, \ldots \mathbf{K}_m)(\alpha_j) \neq 0$ unless for some i $w_j^{(i)} = 0$ for all $\boldsymbol{w}^{(i)} \in \mathbf{V}_{\mathbf{K}^{(i)}}$. It is easy to see however that such a principle cannot hold for any

$\mathfrak{F}$ which satisfies the Consistency Principle. For if we take the example where there are three atoms α_1, α_2, α_3, and $\mathbf{K}_1 = \{\boldsymbol{w}(\alpha_1) = \frac{1}{3}\}$, while $\mathbf{K}_2 = \{\boldsymbol{w}(\alpha_2) = \frac{2}{3}\}$, then by the consistency principle the only possible social belief function is given by $\boldsymbol{v} = < \frac{1}{3}, \frac{2}{3}, 0 >$, despite the fact that neither $\mathbf{K}_1$ nor $\mathbf{K}_2$ on their own force belief in α_3 to be zero. Furthermore, at least in the case of the inference process **SEP**, it is easy to show that similar counterexamples $\mathbf{K}_1^{'}$ and $\mathbf{K}_2^{'}$ to such a principle can be found where the union of the constraint sets $\mathbf{K}_1^{'}$ and $\mathbf{K}_2^{'}$ is not consistent.

Another important property of the inference process **ME**, identified in [28] is that of continuity with respect to the Blaschke topology. At present we do not know whether an analogous formulation of this continuity principle holds for **SEP**, although it seems likely that this is the case.

The **Obstinacy Principle** for an inference process $\mathbf{I}$ states that if $\mathbf{K}$ and $\mathbf{K}^{'}$ are constraint sets such that $\mathbf{I}(\mathbf{K}) \in \mathbf{V}_{\mathbf{K}'}$ then $\mathbf{I}(\mathbf{K} \cup \mathbf{K}^{'}) = \mathbf{I}(\mathbf{K})$. While this principle is satisfied by **ME**, and indeed by nearly all standard inference processes, an appropriate generalisation of the principle to the context of a social inference process is not immediately evident. However Savage in [33] formulates an interesting geometric version of obstinacy for a social inference process which he proves holds for the case of **SEP**. However, unlike the situation for the classical notion of obstinacy with respect to **ME**, it is not clear whether Savage's principle could be of practical value in simplifying calculations of **SEP**-values.

The three remaining principles used to characterise **ME**, in the Paris-Vencovská framework as in [28] are all most naturally formulated in the context where the atoms $\alpha_1, \alpha_2, \ldots \alpha_J$ are the atoms of a Boolean algebra generated by a set of propositional variables $p_1 \ldots p_n$, so that $J = 2^n$. These principles are those of **Language Invariance** , **Irrelevant Information** , and **Independence**, and appropriate reformulations of them for social inference processes are studied in a forthcoming paper by Martin Adamčik and the author [1]. It turns out that whereas language invariance generalises in a straightforward manner and holds for **SEP**, analogues of the other two principles either fail or need careful reformulation if they are to hold for **SEP**.

4 An Alternative Characterisation of SEP

A remarkable characteristic of **SEP** is that the use of maximum entropy at the second second stage of the defining process, which is included in order to force the choice of a social belief function to be unique in cases when this would not otherwise hold, can actually be eliminated by insisting that the social inference process satisfy a variant of the axiom of proportionality. Such an argument counters a possible objection that the invocation of maximum entropy at the second stage of the definition is somewhat artificial. To be precise it is possible to substitute the following procedure to define **SEP**. We will explain and justify the procedure heuristically before formally stating and proving the corresponding theorem.

In order to calculate a unique social belief function $\boldsymbol{v}$ for a college $\mathbf{M}$ with vector of constraint sets $\overrightarrow{\mathbf{K}} = \mathbf{K}_1 \ldots \mathbf{K}_m$, chairman $\mathbf{A}_0$ recognizes that he may have to use a casting constraint set of his own in order to eliminate ambiguities caused by the failure of the agreed process of minimising the sum of cross-entropies to provide a *unique* social belief function. However, as a good chairman, he wishes to intervene in a manner which (a) demonstrates that he is completely unprejudiced, and (b) reduces to an absolute minimum the effect which his own opinion may have on the outcome. In order to fulfil (a) it seems clear to him that he should choose his casting constraint set $\mathbf{K}_0$ to be a constraint set $\mathbf{I}$ with

$$\mathbf{V_I} = \{< \frac{1}{J}, \frac{1}{J} \ldots \frac{1}{J} >\}$$

His only other possible choice would seem to be to take $\mathbf{K}_0$ to be the empty set of constraints, but by Collegiality this would clearly not resolve any ambiguity. On the other hand chairman $\mathbf{A}_0$ worries that if he simply adds in his constraint set $\mathbf{I}$ as a single extra member of the opinion forming body, he may be exerting more influence than is necessary or appropriate, if other opinions are finely balanced. He therefore resolves to dilute his influence in the following manner. Inspired by the Proportionality Principle, he imagines that, for some large finite number n, each member of the college except himself is replaced by exactly n clones, each clone having exactly the same set of constraints as the member replaced; and to this new college of nm members $\mathbf{A}_0$ adds himself as a single additional member with constraint set $\mathbf{I}$ as above.

The vector of sets of constraints of the members of the new college of $nm + 1$ members now looks as follows:

$$\mathbf{K}_1, \ldots, \mathbf{K}_1,\, \mathbf{K}_2, \ldots, \mathbf{K}_2, \ldots\ldots, \mathbf{K}_m, \ldots, \mathbf{K}_m,\, \mathbf{I}$$

Chairman $\mathbf{A}_0$ notices that since $\mathbf{V_I}$ is a singleton, by 3.8 the result of minimising the sum of cross-entropies subject to these constraint sets will, for any given n always yield a unique social belief function. He reasons that if as $n \to \infty$ the resulting sequence of social belief functions converges to a belief function $\boldsymbol{v}$ then this should be an optimal choice as social belief function since his own influence on the process will surely have become as diluted as possible, thus satisfying his condition (b) above. We will prove in Theorem 4.2 below that not only does this sequence of belief functions converge, but that the resulting limiting belief function $\boldsymbol{v}$ will in fact *always be* $\mathbf{SEP}(\overrightarrow{\mathbf{K}})$. This is true whether or not $\boldsymbol{\Delta}(\overrightarrow{\mathbf{K}})$ is a singleton.

Consequently Chairman $\mathbf{A}_0$ can reason that his use of **ME** in the definition of **SEP** is fully justified by the above heuristic.

In order to state formally and prove this result we introduce the following definition:

Definition 4.1

The Weak Social Entropy Process, **WSEP**, is defined by

$$\mathbf{WSEP}(\vec{\mathbf{K}}\,) \;=\; \begin{cases} \boldsymbol{v} & \text{if} \quad \boldsymbol{\Delta}(\vec{\mathbf{K}}) \;\text{ is the singleton } \{\boldsymbol{v}\}, \\ & \text{undefined otherwise.} \end{cases}$$

□

WSEP is of course not a true social inference process since it is only partially defined. Obviously however $\mathbf{WSEP}(\vec{\mathbf{K}}\,) \;=\; \mathbf{SEP}(\vec{\mathbf{K}}\,)$ whenever the former is defined.

We will denote the constraint sets of the college of $nm+1$ members

$$\mathbf{K}_1, \ldots, \mathbf{K}_1,\, \mathbf{K}_2, \ldots, \mathbf{K}_2, \ldots\ldots, \mathbf{K}_m, \ldots, \mathbf{K}_m,\, \mathbf{I}$$

in abbreviated form by

$$n\vec{\mathbf{K}}\,, \mathbf{I}\,.$$

Theorem 4.2

For any $\vec{\mathbf{K}}$ and any $n \in \mathbb{N}^+$ $\quad \mathbf{WSEP}(n\vec{\mathbf{K}}\,,\mathbf{I}) = \mathbf{SEP}(n\vec{\mathbf{K}}\,,\mathbf{I})$, and furthermore

$$\lim_{n\to\infty} \mathbf{WSEP}(n\vec{\mathbf{K}}\,,\mathbf{I}) \;=\; \mathbf{SEP}(\vec{\mathbf{K}}\,)$$

Proof:

Since $\mathbf{V_I}$ is a singleton, by 3.8 $\boldsymbol{\Delta}(n\vec{\mathbf{K}},\mathbf{I})$ is always a singleton, and so $\mathbf{WSEP}(n\vec{\mathbf{K}}\,,\mathbf{I})$ is a well-defined point for any $n \in \mathbb{N}^+$. It does not follow from this that $\boldsymbol{\Gamma}(n\vec{\mathbf{K}},\mathbf{I})$ is a singleton, but we will show below that "significant" coordinates are uniquely determined.

For now let us fix n. Then if $\mathbf{WSEP}(n\vec{\mathbf{K}}\,,\mathbf{I}) \;=\; \boldsymbol{v}$ say, and noting that $\mathrm{Sig}_{n\vec{\mathbf{K}},\mathbf{I}} \;=\; \mathrm{Sig}_{\vec{\mathbf{K}}}$, then for every $j \in J$

$$v_j \;=\; 0 \quad \text{if and only if} \quad j \notin \mathrm{Sig}_{\vec{\mathbf{K}}}$$

This is true because if $\vec{\boldsymbol{w}}$ is a point in $\boldsymbol{\Gamma}(n\vec{\mathbf{K}},\mathbf{I})$ which generates $\boldsymbol{v}$, then if $j \in \mathrm{Sig}_{\vec{\mathbf{K}}}$ then since $w_j^{(mn+1)} \;=\; \frac{1}{J}$ it follows from 3.6(i) that every entry in the column vector $\boldsymbol{w}_j$ is non-zero, so v_j is non-zero.
Furthermore for any such $\vec{\boldsymbol{w}}$ in $\boldsymbol{\Gamma}(n\vec{\mathbf{K}},\mathbf{I})$ it is clear that the first n rows, i.e. with $i \;=\; 1\ldots n$, which correspond to the members with constraint set $\mathbf{K_1}$, must all be identical for those entries $w_j^{(i)}$ with $j \in \mathrm{Sig}_{\vec{\mathbf{K}}}$. For if two of these rows were not so identical then, if they differed in the j'th entry for some $j \in \mathrm{Sig}_{\vec{\mathbf{K}}}$, we could interchange them to obtain a different point $\overrightarrow{\boldsymbol{w'}}$ in $\boldsymbol{\Gamma}(n\vec{\mathbf{K}},\mathbf{I})$: however the j'th column $\boldsymbol{w'}_j$ could not then be a multiple of $\boldsymbol{w}_j$, contradicting Theorem 3.6.

Moreover exactly the same argument works for the second and subsequent blocks of n rows, up to the m'th block of n rows.

From the above observations it follows that finding an $\vec{\boldsymbol{w}}$ in $\mathbf{\Gamma}(n\vec{\mathbf{K}},\mathbf{I})$ is essentially the same problem as finding an $\vec{\boldsymbol{x}} \in \bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}$ for which

$$\sum_{j=1}^{J}\left[\frac{1}{J}\prod_{i=1}^{m}(x_j^{(i)})^n\right]^{\frac{1}{nm+1}} \quad \text{is maximal,}$$

or equivalently, for which the function defined by

$$\mathbf{H}_{\epsilon(n)}(\vec{\boldsymbol{x}}\,) \;=\; \sum_{j=1}^{J}\left[\left[\prod_{i=1}^{m}x_j^{(i)}\right]^{\frac{1}{m}}\right]^{(1-\epsilon(n))} \qquad \text{is maximal,}$$

where $\epsilon(n) \;=\; \frac{1}{mn+1} \;\to\; 0$ as $n \;\to\; \infty$.

Note that for any $\epsilon(n)$ as above the values of $\left[\prod_{i=1}^{m} x_j^{(i)}\right]^{\frac{1}{m}}$ for which $\mathbf{H}_{\epsilon(n)}(\vec{\boldsymbol{x}}\,)$ is maximal are uniquely determined for each $j \;=\; 1\ldots J$ and are non-zero if and only if $j \in \mathrm{Sig}_{\vec{\mathbf{K}}}$.

In order to make what follows more readable, we shall temporarily write ϵ instead of $\epsilon(n)$ and suppress the dependence of ϵ on n.

For any such ϵ as above we denote the vector of unique values of $\left[\prod_{i=1}^{m} x_j^{(i)}\right]^{\frac{1}{m}}$ as defined above by

$$\boldsymbol{y}_\epsilon \;=\; < y_{1,\,\epsilon}\ldots y_{J,\,\epsilon} > \tag{25}$$

and we denote the maximal value of $\mathbf{H}_\epsilon(\vec{\boldsymbol{x}}\,)$ by m_ϵ , so that

$$m_\epsilon \;=\; \sum_{j=1}^{J}(y_{j,\,\epsilon})^{1-\epsilon} \tag{26}$$

and let

$$M_\epsilon \;=\; \sum_{j=1}^{J} y_{j,\,\epsilon} \tag{27}$$

We need to examine the behaviour of $\boldsymbol{y}_\epsilon$ as $\epsilon \to 0$, i.e. as $n \to \infty$.

Define M_0 to be $M_{\vec{\mathbf{K}}}$, i.e the maximum possible value of $\sum_{j=1}^{J}\left[\prod_{i=1}^{m} x_j^{(i)}\right]^{\frac{1}{m}}$. By our initial assumptions $M_0 > 0$.

A straight forward consequence of the above definitions is the following:

Lemma 4.3

$$M_\epsilon \;\leq\; M_0 \;\leq\; m_\epsilon \quad \text{for all} \quad \epsilon \in (0,1)$$

□

Lemma 4.4

$$M_\epsilon \to M_0 \quad \text{as} \quad \epsilon \to 0^+ \,.$$

Proof:

We show first that the function $y^{1-\epsilon}$ converges uniformly to y as $\epsilon \to 0^+$ in the sense that there exists some positive real valued function $T(\epsilon)$ such that for all $y \in [0,1]$ and all ϵ with $0 < \epsilon < \frac{1}{2}$

$$y^{1-\epsilon} \;<\; y \,+\, T(\epsilon) \quad \text{ and } \quad lim_{\epsilon \to 0^+}\, T(\epsilon) \;=\; 0\,.$$

Now

$$y^{1-\epsilon} \;=\; y\, e^{-\epsilon \log y}$$

whence , expanding the exponential function as a power series, multiplying by y, and taking out a factor of ϵ, we get

$$y^{1-\epsilon} \,-\, y \;=\; \epsilon \sum_{k=1}^{\infty} \frac{\epsilon^{k-1}\, y\,(-\log y)^k}{k!}$$

The absolute value of $y\,(\log y)^k$ is at a maximum when $y \,=\, e^{-k}$ and hence the absolute value of the k'th term of the above series is bounded by $\frac{\epsilon^k}{k!}\left[\frac{k}{e}\right]^k$. Since this bound decreases for decreasing ϵ, we have that for all ϵ with $0 < \epsilon < \frac{1}{2}$ and $y \in [0,1]$

$$y^{1-\epsilon} \,-\, y \;<\; \epsilon \sum_{k=1}^{\infty} \left[\frac{1}{2}\right]^{k-1} \left[\frac{k}{e}\right]^{k} \frac{1}{k!}$$

and since the sum converges by d'Alembert's ratio test, the right hand side provides the required function $T(\epsilon)$.

To complete the proof of 4.4 we note that, using 4.3 and the above,

$$M_\epsilon \;\le\; M_0 \;\le\; m_\epsilon \;=\; \sum_{j=1}^{\infty} (y_{j,\,\epsilon})^{1-\epsilon} \;\le\; \sum_{j=1}^{\infty} y_{j,\,\epsilon} \,+\, T(\epsilon) \;=\; M_\epsilon \,+\, T(\epsilon)\,.$$

Hence , letting ϵ tend to zero, we obtain the required result.

$\square$

We now note that for fixed ϵ an equivalent definition of $\boldsymbol{y}_\epsilon$ is as that vector of values which maximises the function

$$\mathbf{G}_\epsilon(\boldsymbol{y}) \;=\; \frac{1}{\epsilon}\log\left[\frac{\sum_{j\in \mathrm{Sig}_{\vec{\mathbf{K}}}} (y_j)^{(1-\epsilon)}}{M_\epsilon}\right] \tag{28}$$

subject to the conditions that

$$y_j \;=\; \left[\prod_{i=1}^{m} (x_j^{(i)})\right]^{\frac{1}{m}} \quad \text{for } j \in \mathrm{Sig}_{\vec{\mathbf{K}}}\text{ , and } \;\vec{\boldsymbol{x}} \;\in\; \bigotimes_{i=1}^{m} \mathbf{V}_{\mathbf{K}_i}\;. \tag{29}$$

For fixed ϵ we will now consider the behaviour of $\mathbf{G}_\epsilon(\boldsymbol{y})$ for general $\boldsymbol{y}$ satisfying conditions (29) above. Actually we are only interested in those $\boldsymbol{y}$ which are either of the form $\boldsymbol{y}_{\,\epsilon(\mathrm{n})}$ for some n or which are such that $\sum_{j=1}^{J} y_j \;=\; M_0$, and from now on we shall assume that $\boldsymbol{y}$ is of this kind. We note that for $j \in \mathrm{Sig}_{\vec{\mathbf{K}}}$ $0 < y_j \le 1$, and that for such y_j for any $k \in \mathbb{N}^+$ $|y_j(\log y_j)^k|$ is uniformly bounded

above by $(\frac{k}{e})^k$, (as in the proof of 4.4).
By 4.4 it follows that

$$J \;\geq\; \sum_{j=1}^{J} y_j \;>\; c \;>\; 0 \tag{30}$$

for some fixed bound c for all such $\boldsymbol{y}$.

Now

$$\begin{aligned} (y_j)^{1-\epsilon} &= (y_j)\, e^{-\epsilon \log y_j} && (31)\\ &= y_j - \epsilon\, y_j \log y_j \;+ \sum_{k=2}^{\infty} y_j\, (-\epsilon \log y_j)^k && (32) \end{aligned}$$

whence

$$\sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} (y_j)^{1-\epsilon} \;=\; \sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} y_j \left[1 - \epsilon \frac{\sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} \; y_j \log y_j}{\sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} \; y_j} + \mathrm{O}(\epsilon^2) \right] \tag{33}$$

where the term $\mathrm{O}(\epsilon^2)$ is such that its modulus is, by the argument in the proof of 4.4, uniformly bounded by ϵ^2D for some positive constant D.
Rewriting the equation (4) we now have

$$\mathbf{G}_\epsilon(\boldsymbol{y}) \;=\; \frac{1}{\epsilon} \left[\log \left[1 - \epsilon \frac{\sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} \; y_j \log y_j}{\sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} \; y_j} + \mathrm{O}(\epsilon^2) \right] + \log \frac{\sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} \; y_j}{M_\epsilon} \right] \tag{34}$$

Expanding the logarithm as a power series and using (6) we obtain

$$\mathbf{G}_\epsilon(\boldsymbol{y}) \;=\; -\frac{\sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} \; y_j \log y_j}{\sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} \; y_j} + \epsilon\, \mathrm{R}(\epsilon, \boldsymbol{y}) + \frac{1}{\epsilon} \log \frac{\sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} \; y_j}{M_\epsilon} \tag{35}$$

where $\mid \mathrm{R}(\epsilon, \boldsymbol{y}) \mid$ has a uniform bound independent of $\boldsymbol{y}$ and of ϵ.

Now notice the following facts about equation (35):

1. For given $\epsilon \;=\; \epsilon(n)$ corresponding to a specific value of n, the vector $\boldsymbol{y}_\epsilon$ satisfies

$$\mathbf{G}_\epsilon(\boldsymbol{y}_\epsilon\,) \;=\; -\frac{\sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} \; y_{j,\,\epsilon} \log y_{j,\,\epsilon}}{M_\epsilon} + \epsilon\, \mathrm{R}(\epsilon, \boldsymbol{y}_\epsilon\,) \tag{36}$$

   since the final term vanishes.

2. For any $\boldsymbol{y}$ for which $\sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} \; y_j \;=\; M_0$ the final term of (35) is positive since $M_\epsilon \;\leq\; M_0$ by 4.3.

Let us denote by $\boldsymbol{z}$ that unique $\boldsymbol{y}$ for which

$$\sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} y_j \;=\; M_0 \qquad \text{and} \qquad \sum_{j \in \mathrm{Sig}_{\vec{\mathbf{K}}}} -y_j \log y_j \quad \text{is maximal.}$$

Then

$$\mathbf{SEP}(\vec{\mathbf{K}}\,) \;=\; < \frac{z_1}{M_0} \cdots \frac{z_{\mathrm{J}}}{M_0} > \tag{37}$$

since

$$\sum_{j\in \mathrm{Sig}_{\vec{K}}} -y_j \log y_j \quad \text{is maximal if and only if} \quad \sum_{j\in \mathrm{Sig}_{\vec{K}}} -\frac{y_j}{M_0}\log\frac{y_j}{M_0} \quad \text{is maximal.}$$

To complete the proof of the theorem we need to show that

$$\lim_{n\to\infty} \;<\; \frac{y_{1,\,\epsilon(n)}}{M_{\epsilon(n)}}\dots\frac{y_{\mathrm{J},\,\epsilon(n)}}{M_{\epsilon(n)}} \;>\; = \;<\; \frac{z_1}{M_0}\dots\frac{z_{\mathrm{J}}}{M_0} \;> \tag{38}$$

Since by 4.4 $M_\epsilon \to M_0$ as $\epsilon \to 0$, it suffices to show that $\boldsymbol{y}_\epsilon \to \boldsymbol{z}$ as $\epsilon \to 0$.

Now since all the $\boldsymbol{y}$ are in $[0,1]^J$, by compactness the sequence of $\boldsymbol{y}_{\epsilon(n)}$ for $n \in \mathbb{N}$ has a convergent subsequence, say $\boldsymbol{y}_{\epsilon(\rho(n))}$, where $\epsilon(\rho(n)) \to 0$ as $n \to \infty$.
Let

$$\lim_{n\to\infty} \boldsymbol{y}_{\epsilon(\rho(n))} \;=\; \boldsymbol{y}^* \,. \tag{39}$$

Then from (36) above and the fact that $M_\epsilon \to M_0$, it follows that

$$\lim_{n\to\infty} \mathbf{G}_{\epsilon(\rho(n))}(\boldsymbol{y}_{\epsilon(\rho(n))}) \;=\; -\frac{\sum_{j\in \mathrm{Sig}_{\vec{K}}} y_j^* \log y_j^*}{M_0} \tag{40}$$

and

$$\sum_{j\in \mathrm{Sig}_{\vec{K}}} y_j^* \;=\; M_0 \tag{41}$$

We now show that $\boldsymbol{y}^* \;=\; \boldsymbol{z}$.

For suppose for contradiction that this were not so. Let

$$\frac{1}{M_0}\left[-\sum_{j\in \mathrm{Sig}_{\vec{K}}} z_j \log z_j \;+\; \sum_{j\in \mathrm{Sig}_{\vec{K}}} y_j^* \log y_j^*\right] \;=\; d \tag{42}$$

Then $d > 0$ since $\boldsymbol{y}^*$ and $\boldsymbol{z}$ both have sum M_0 and $\boldsymbol{z}$ is the unique maximum entropy point. Now by (35)

$$\begin{aligned}
\mathbf{G}_{\epsilon(\rho(n))}(\boldsymbol{z}) \quad &= \quad d - \frac{\sum_{j\in \mathrm{Sig}_{\vec{K}}} y_j^* \log y_j^*}{M_0} + \epsilon(\rho(n))\,\mathrm{R}(\epsilon(\rho(n)),\boldsymbol{z}) + \frac{1}{\epsilon(\rho(n))}\log\frac{M_0}{M_{\epsilon(\rho(n))}} \\
&\geq \quad d - \frac{\sum_{j\in \mathrm{Sig}_{\vec{K}}} y_j^* \log y_j^*}{M_0} + \epsilon(\rho(n))\,\mathrm{R}(\epsilon(\rho(n)),\boldsymbol{z})
\end{aligned} \tag{43}$$

However for large enough the right hand-side is strictly greater than $\mathbf{G}_{\epsilon(\rho(n))}(\boldsymbol{y}_{\epsilon(\rho(n))})$ by (36),(40), and the boundedness of R.
This is impossible since then $\mathbf{G}_{\epsilon(\rho(n))}(\boldsymbol{z}) > \mathbf{G}_{\epsilon(\rho(n))}(\boldsymbol{y}_{\epsilon(\rho(n))})$ which contradicts the definition of $\boldsymbol{y}_{\epsilon(\rho(n))}$. Thus we have shown that $\boldsymbol{y}^* = \boldsymbol{z}$.
It remains to show that the whole sequence of the $\boldsymbol{y}_{\epsilon(n)}$ converges to $\boldsymbol{z}$ as $n \to \infty$. If this were not the case then there would be some $\delta > 0$ such that there exists an infinite subsequence $\boldsymbol{y}_{\epsilon(\tau(n))}$ of the $\boldsymbol{y}_{\epsilon(n)}$ such that the $\boldsymbol{y}_{\epsilon(\tau(n))}$ are bounded away from $\boldsymbol{z}$ by Euclidean distance $\mid \boldsymbol{y}_{\epsilon(\tau(n))} - \boldsymbol{z} \mid > \delta$ for all $n \in \mathbb{N}$. However now by compactness again this subsequence $\boldsymbol{y}_{\epsilon(\tau(n))}$ itself has an infinite convergent subsequence which converges to a point say $\boldsymbol{y}^{**}$. By the same argument as for $\boldsymbol{y}^*$ we must have that $\boldsymbol{y}^{**} = \boldsymbol{z}$; on the other hand by its definition $\boldsymbol{y}^{**}$ is bounded away from $\boldsymbol{z}$ by distance at least δ, which gives a contradiction. Thus we have

established (38) and the proof of Theorem 4.2 is complete.

□

It is perhaps worth remarking that in the very special case when there is only a single member $\mathbf{A}_1$ of the college apart from the Chairman $\mathbf{A}_0$, the explanation of Theorem 4.2 given at the beginning of this section provides a new interpretation of an old technical result. For in this special case, for any $n \in \mathbb{N}$
$\mathbf{WSEP}(n\mathbf{K}_1, \mathbf{I})$ returns that probability function $\boldsymbol{v}$ which satisfies the constraints $\mathbf{K}_1$ and which maximises the function

$$\sum_{j=1}^{J} v_j^{(\frac{n}{n+1})}$$

In other words for a given $n \in \mathbb{N}$ this gives the same result as applying the Renyi inference process $\mathbf{REN}_r$ with parameter $r = (\frac{n}{n+1})$. Now it is an old result (see e.g. [23] or [14]) that as r $\rightarrow$ 1 the result of applying the Renyi process $\mathbf{REN}_r$ to a given set of constraints $\mathbf{K}_1$ tends to the maximum entropy solution for $\mathbf{K}_1$. So the heuristic explanation underlying Theorem 4.2 may be regarded as a generalised interpretation of this classical result from a new perspective.

References

[1] M.Adamčik and G.M.Wilmers, *Some Invariance Properties of the Social Entropy Process*, in preparation.

[2] S.Boyd and L.Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[3] Imre Csiszár, *Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems*, Annals of Statistics, 19(4) pp. 2032-2066, 1991

[4] Imre Csiszár, *Axiomatic Characterisations of Information Measures*, Entropy, 10, pp.261-273, 2008

[5] R.M.Cooke, *Experts in Uncertainty: Opinion and Subjective Probability*, in Science. Environmental Ethics and Science Policy Series, Oxford University Press, New York, 1991.

[6] D.K. Fadeev, *Zum Begriff der Entropie einer endlichen Wahrscheinlichkeitsschemas*, Arbeiten zur Informationstheorie I, Berlin, Deutscher Verlag der Wissenschaften, 1957, pp 8590.

[7] Simon French, *Group Consensus Probability Distributions: A Critical Survey*, in J. M. Bernardo, M. H. De Groot, D. V. Lindley, and A. F. M. Smith (Eds.), Bayesian Statistics, Elsevier, North Holland, 1985, pp. 183-201.

[8] Ashutosh Garg, T. S. Jayram, Shivakumar Vaithyanathan, Huaiyu Zhu, *Generalized Opinion Pooling*, in Proceedings of the 8th Intl. Symp. on Artificial Intelligence and Mathematics, 2004.

[9] C.Genest, *A characterization theorem for externally Bayesian groups*, Annals of Statistics 12(3) pp. 11001105.

[10] C.Genest, *A conflict between two axioms for combining subjective distributions.* J. Royal Statistical Society, 46(3)pp. 403-405, 1984.

[11] C.Genest and C.G.Wagner, *Further evidence against independence preservation in expert judgement synthesis.* Aequationes Mathematicae 32(1) pp. 74-86, 1987.

[12] C.Genest and J.V.Zidek, *Combining probability distributions: A critique and an annotated bibliography.* Statistical Science Vol 1 No 1 pp. 114-135, 1986.

[13] C.Genest,K.J.McConway,M.J.Schervish, *Characterization of externally Bayesian pooling operators*, Ann. Statist. 14(2) pp. 487-501, 1986.

[14] Peter Hawes, *An Investigation of Properties of Some Inference Processes*, PhD Thesis, Manchester University, MIMS eprints, 2007, available from http://eprints.ma.man.ac.uk/1304/

[15] E.T. Jaynes, *Where do we Stand on Maximum Entropy?* in "The Maximum Entropy Formalism", R. D. Levine and M. Tribus (eds.), M. I. T. Press, Cambridge, MA, , 1979.

[16] E.T. Jaynes, *The Well-Posed Problem*, Foundations of Physics, vol. 3, No.4, 1973.

[17] Jan Kracík , *On composition of probability density functions* , in "Multiple Participant Decision Making", pp. 113-121, Workshop on Computer-Intensive Methods in Control and Data Processing 2004, Prague.

[18] Jan Kracík, *Cooperation Methods in Bayesian Decision Making with Multiple Participants*, Ph.D. Thesis, Czech Technical University, Prague, February 2009.

[19] S. Kullback and R. Leibler, *On information and sufficiency*, Annals of Mathematical Statistics, 22, pp. 79-87, 1951.

[20] S.Kullback, *Information Theory and Statistics*, Wiley, New York, 1959.

[21] W.B.Levy and H.Delic, *Maximum entropy aggregation of individual opinions.* IEEE Trans. Systems, Man, Cybernetics 24(4) pp. 606-613, 1994.

[22] Albert Madansky, *Externally Bayesian Groups*, Technical Report RM-4141-PR, RAND Corporation, 1964.

[23] Ismail Mohamed , MSc dissertation, Manchester University, 1998.

[24] Jae Myung, Sridhar Ramamoorti, A.D.Bailey,Jr., *Maximum Entropy Aggregation of Expert Predictions*, in Management Science, Vol. 42, No. 10 , pp. 1420-1436, 1996.

[25] Daniel Osherson and Moshe Vardi, *Aggregating disparate estimates of chance, Games and Economic Behavior*, pp. 148-173, July 2006.

[26] J B Paris and A. Vencovská, *On the applicability of maximum entropy to inexact reasoning*, International Journal of Approximate Reasoning, vol. 3(1),pp.1-34, 1989.

[27] J.B.Paris and Alena Vencovská, *A Note on the Inevitability of Maximum Entropy*, International Journal of Approximate Reasoning, 4, pp. 183-224, 1990.

[28] J.B.Paris, *The Uncertain Reasoner's Companion - A Mathematical Perspective*, Cambridge University Press, Cambridge, UK, 1994.

[29] J.B.Paris and Alena Vencovská, *In defence of the Maximum Entropy Inference Process.* International Journal of Approximate Reasoning, vol.17, no.1, pp. 77-103, 1997.

[30] J.B.Paris, *Common sense and maximum entropy.* Synthese, vol.117, pp. 75-93, 1999.

[31] A. Renyi *On measures of entropy and information*, Proceedings of the 4th Berkeley Symposium in Mathematical Statistics, 1960.

[32] A. Renyi *Wahrscheinlichkeitsrechnung, Berlin: Deutscher Verlag der Wissenschaften*, 1962.

[33] S.D.Savage, *The Logical and Philosophical Foundations of Social Inference Processes*, MSc Dissertation, University of Manchester, 2010.

[34] C.E. Shannon and W.Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.

[35] J.E.Shore and R.W.Johnson, *Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy*, IEEE Transactions on Information Theory, IT-26(1)pp. 26-37, 1980.

[36] C.Wagner, *Aggregating subjective probabilities: Some limitative theorems.* Notre Dame J. Formal Logic 25(3) pp. 233-240 , 1984.

[37] T.S.Wallsten, D.V.Budescu, Ido Erev, and Adele Diederich, *Evaluating and Combining Subjective Probability Estimates*, Journal of Behavioral Decision Making, Vol. 10, 1997.

[38] George Wilmers, *The Social Entropy Process: Axiomatising the Aggregation of Probabilistic Beliefs*, in Probability, Uncertainty and Rationality, pp 87-104, edited by Hykel Hosni and Franco Montagna, 10 CRM series, Scuola Normale Superiore, Pisa, 2010.