

*The Social Entropy Process: Axiomatising the
Aggregation of Probabilistic Beliefs*

Wilmers, George

2010

MIMS EPrint: **2010.13**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: <http://eprints.maths.manchester.ac.uk/>

And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

The Social Entropy Process: Axiomatising the Aggregation of Probabilistic Beliefs

by

George Wilmers

1. Introduction

The present work stems from a desire to combine ideas arising from two historically different schemes of probabilistic reasoning, each having its own axiomatic traditions, into a single broader axiomatic framework, capable of providing general new insights into the nature of probabilistic inference in a multiagent context. In the present sketch of our work we first describe briefly the background context, and we then 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. Finally 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 the maximum entropy inference process: indeed it can be regarded as a natural extension of that inference process to the multiagent context.

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 agents¹.

Proofs of the results in the present paper, while reasonably straightforward, have mostly not been included, but will be included in a more detailed version of this

¹This condition is sometimes known as the Watts Assumption (see [12]).

work which will appear elsewhere.

In order to fix notation let $S = \{\alpha_1, \alpha_2, \dots, \alpha_J\}$ denote some fixed finite set of mutually exclusive and exhaustive outcomes or atomic events. A probability function w on S is a function $w : S \rightarrow [0, 1]$ such that $\sum_{j=1}^J w(\alpha_j) = 1$. Slightly abusing notation we will identify w with the vector of values $\langle w_1 \dots w_J \rangle$ where w_j denotes $w(\alpha_j)$ for $j = 1 \dots J$. If such a w represents the subjective belief of an individual \mathbf{A} in the outcomes of S we refer to w as \mathbf{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 w is assumed to extend so as to take values on complex events in the standard way, i.e. for any θ

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

where \vDash denotes the classical notion of logical implication. Conditional probabilities are defined in the usual manner. We note that in this paper the use the term “belief function” will always denote a probability function in the above sense.

The first scheme referred to above is the notion of an *inference process* first formulated by Paris and Vencovská some twenty years ago (see [11], [12], [13], and [14]). The problematic of Paris and Vencovská is that of an isolated individual \mathbf{A} whose belief function is in general not completely specified, but whose set of beliefs is instead regarded as a set of constraints \mathbf{K} on the possible values which the vector $\langle w_1 \dots w_J \rangle$ may take. The constraint set \mathbf{K} therefore defines a certain region of the Euclidean space \mathbb{R}^J , denoted by $\mathbf{V}_{\mathbf{K}}$, consisting of all vectors $\langle w_1 \dots w_J \rangle$ which satisfy the constraints in \mathbf{K} together with the conditions that $\sum_{j=1}^J w_j = 1$ and that $w_j \geq 0$ for all j . It is assumed that the constraint sets \mathbf{K} which we consider are consistent (i.e. $\mathbf{V}_{\mathbf{K}}$ is non-empty), and are such that $\mathbf{V}_{\mathbf{K}}$ has pleasant geometrical properties. More precisely, the exact requirement on a set of constraints \mathbf{K} is that the set $\mathbf{V}_{\mathbf{K}}$ 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 sets². Paris and Vencovská ask the question: given any such \mathbf{K} , by what rational principles should \mathbf{A} choose his probabilistic belief function w consistent with \mathbf{K} *in the absence of any other information?*

²This formulation ensures that conditions such as $w(\theta) = \frac{1}{3}$, $w(\phi | \psi) = \frac{4}{5}$, and $w(\psi | \theta) \leq \frac{1}{2}$, where θ, ϕ , and ψ are Boolean combinations of the α_j 's, are all permissible in \mathbf{K} provided that they are consistent. Here a conditional constraint such as $w(\psi | \theta) \leq \frac{1}{2}$ is interpreted as $w(\psi \wedge \theta) \leq \frac{1}{2} w(\theta)$ which is always a well-defined linear constraint, albeit trivial when $w(\theta) = 0$.

A set of constraints \mathbf{K} as above is often, slightly misleadingly, called a knowledge base. A rule \mathbf{I} which for every such \mathbf{K} chooses such a $w \in \mathbf{V}_{\mathbf{K}}$ is called an *inference process*. Given \mathbf{K} we denote the belief function w chosen by \mathbf{I} by $\mathbf{I}(\mathbf{K})$. The question above can then be reformulated as: what self-evident general principles should an inference process \mathbf{I} satisfy? This question has been intensively studied over the last twenty years and much is known. In particular in [11] Paris and Vencovská found an elegant set of principles which uniquely characterise the maximum entropy inference process³, \mathbf{ME} , which is defined as follows: given \mathbf{K} as above, $\mathbf{ME}(\mathbf{K})$ chooses that unique probability distribution w which maximises the Shannon entropy of w ,

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

subject to the condition that $w \in \mathbf{V}_{\mathbf{K}}$. Although some of the principles used to characterise \mathbf{ME} may individually be open to philosophical challenge, they are sufficiently convincing overall to give \mathbf{ME} the appearance of a gold standard, in the sense that no other known inference process satisfies an equally convincing set of principles⁴.

An apparently rather different problematic of probabilistic inference has been much studied in decision theoretic literature. Given possible outcomes $\alpha_1, \alpha_2, \dots, \alpha_J$ as before, let $\{\mathbf{A}_i \mid i = 1 \dots m\}$ be a finite set of agents each of whom possesses his own particular probabilistic belief function $w^{(i)}$ on the set of outcomes, and let us suppose that these $w^{(i)}$ have already been determined. How then should these individual belief functions be aggregated so as to yield a single probabilistic belief function 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 \mathbf{LinOp} and \mathbf{LogOp} . \mathbf{LinOp} takes v to be the arithmetic mean of the $w^{(i)}$, i.e.

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

³This characterisation considerably strengthens earlier work of Shore and Johnson in [16].

⁴Other favored inference processes which satisfy many, but not all, of these principles are the minimum distance inference process, \mathbf{MD} , the limit centre of mass process, \mathbf{CM}^∞ , all \mathbf{Renyi} inference processes, and the remarkable $\mathbf{Maximin}$ process of Hawes [8]. (See Paris [12] for a general introduction to inference processes, and also Hawes [8], especially the comparative table in Chapter 9, for an excellent résumé of the current state of knowledge concerning this topic).

whereas **LogOp** chooses v to be the normalised geometric mean given by:

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

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 e.g. [6], [2], [1], [4], [5], [7], [15] 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 the framework of pooling operators as a very special case. Thus we consider, for any $m \geq 1$, a set \mathbf{M} consisting of m individuals $\mathbf{A}_1 \dots \mathbf{A}_m$, each of whom possesses his own set of consistent constraints, respectively $\mathbf{K}_1 \dots \mathbf{K}_m$, on his possible belief function on the set of outcomes $\alpha_1, \alpha_2, \dots \alpha_J$. (Note that we are only assuming 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*.

A *social inference process*, \mathfrak{F} , is a function which chooses, for any such $m \geq 1$ and $\mathbf{K}_1 \dots \mathbf{K}_m$, a probability function on $\alpha_1, \alpha_2, \dots \alpha_J$, denoted by $\mathfrak{F}(\mathbf{K}_1 \dots \mathbf{K}_m)$, which we refer to as the *social belief function* defined by $\mathfrak{F}(\mathbf{K}_1 \dots \mathbf{K}_m)$. Note that, trivially, provided that when $m = 1$ $\mathfrak{F}(\mathbf{K}_1) \in \mathbf{V}_{\mathbf{K}_1}$ for all \mathbf{K}_1 , \mathfrak{F} marginalises to an inference process. On the other hand, in the special case where $\mathbf{V}_{\mathbf{K}_i}$ is a singleton for all $i = 1 \dots m$, \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 uncertainty⁵.

2. An Axiomatic Framework for a Social Inference Process

The underlying idea of a social inference process is not new. (See e.g. [17]). 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 $w^{(i)}$ appropriate to that agent, and a pooling operator is then chosen and applied to the set of $w^{(i)}$ to yield a social belief function⁶. 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 however completely different. We reject the two stage approach above on the grounds that the classical notion of an inference process applies to

⁵The interested reader may consult [13] for a detailed analysis of this point in connection with supposed paradoxes arising from "representation dependence".

⁶We note that by no means are all authors reductionist in this sense: in particular although their concerns are somewhat different from ours, neither [17] nor [9] make such an assumption.

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 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 w it is clear that certain "information" has been discarded⁷; thus, while w is determined by \mathbf{K} once an inference process is given, neither \mathbf{K} nor $\mathbf{V}_{\mathbf{K}}$ can be recaptured from 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 $\mathbf{V}_{\mathbf{K}} = \{ \langle \frac{1}{j}, \frac{1}{j} \dots \frac{1}{j} \rangle \}$, although the application of \mathbf{ME} , or of any other reasonable inference process, yields $w = \langle \frac{1}{j}, \frac{1}{j} \dots \frac{1}{j} \rangle$ 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 $w^{(i)}$ so as to take account of the information provided by the other individuals, in such a way that $w^{(i)}$ 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 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 $w^{(i)}$ which is attributed to him by the procedure above: such an $w^{(i)}$ 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 $w^{(i)}$ has an intersubjective or normative character: however the calculation of $w^{(i)}$ now depends not only on \mathbf{K}_i but on the belief sets of all the other members of the college.

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

Considerations similar to the above give rise to the following radical but attractive principle for a social inference process to satisfy:

The Collegial Principle

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

$$\mathfrak{F}(\mathbf{K}_1 \dots \mathbf{K}_m) = \mathfrak{F}(\mathbf{K}_1 \dots \mathbf{K}_k)$$

□

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.

We now introduce a number of other desirable principles for a social inference process \mathfrak{F} to satisfy. Several of 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 $i = 1 \dots m$ $\mathbf{V}_{\mathbf{K}_i} = \mathbf{V}_{\mathbf{K}'_i}$ then

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

.

□

Otherwise expressed the Equivalence Principle states that substituting constraint sets which are equivalent in the sense that the set of probability 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. [12]). 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 Δ is a non-empty closed convex set of belief functions then we may write $\mathbf{ME}(\Delta)$ to denote the unique $w \in \Delta$ which maximises the Shannon entropy function.

The Anonymity Principle

This principle states that $\mathfrak{F}(\mathbf{K}_1 \dots \mathbf{K}_m)$ depends only on the multiset of constraint sets $\{\mathbf{K}_1 \dots \mathbf{K}_m\}$ 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. □

In order to ensure that \mathfrak{F} behaves like an inference process for the case $m = 1$ we need the following axiom

The Consistency Axiom

For the case when $m = 1$

$$\mathfrak{F}(\mathbf{K}_1) \in \mathbf{V}_{\mathbf{K}_1}$$

for any constraint set \mathbf{K}_1 . □

It is immediate from the Consistency, Anonymity, and Collegiality that \mathfrak{F} satisfies the “Unanimity” property that for any \mathbf{K} ,

$$\mathfrak{F}(\mathbf{K} \dots \mathbf{K}) = \mathfrak{F}(\mathbf{K}).$$
□

Another immediate consequence of the above axioms is:

Lemma 2.1

If \mathfrak{F} satisfies Consistency and Collegiality, then if $\mathbf{K}_1 \dots \mathbf{K}_m$ are such that

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

then

$$\mathfrak{F}(\mathbf{K}_1 \dots \mathbf{K}_m) \in \bigcap_{i=1}^m \mathbf{V}_{\mathbf{K}_i}$$
□

The following principle is again a familiar one satisfied by classical inference processes (cf. [12]):

Let σ denote a permutation of the atoms of S . Such a σ induces a corresponding permutation on the coordinates of probability distributions $\langle w_1 \dots w_J \rangle$, 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 $\mathbf{K}_1 \dots \mathbf{K}_m$

$$\mathfrak{F}(\sigma(\mathbf{K}_1) \dots \sigma(\mathbf{K}_m)) = \sigma(\mathfrak{F}(\mathbf{K}_1 \dots \mathbf{K}_m))$$

.

□

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 w denote \mathbf{A}_1 's belief function. Since we are considering the case when $m = 1$ we will drop the superscript from $w^{(1)}$ for ease of notation. For some non-empty set of atoms $\{\alpha_{j_1} \dots \alpha_{j_t}\}$ let ϕ denote the event $\bigvee_{r=1}^t \alpha_{j_r}$. Suppose that \mathbf{K} denotes a set of constraints on the variables $w_{j_1} \dots w_{j_t}$ 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* .

Now let \hat{w}_r denote $w(\alpha_{j_r} \mid \phi)$ for $r = 1 \dots t$, with the \hat{w}_r undefined if $w(\phi) = 0$. Then $\hat{w} = \langle \hat{w}_1 \dots \hat{w}_t \rangle$ is a probability distribution provided that $w(\phi) \neq 0$. Let \mathbf{K} be a nice set of constraints on the probability distribution \hat{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 $w(\phi) = 0$.

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)

If \mathbf{K}_1 is a nice set of constraints *conditioned on* ϕ , and \mathbf{K}_1^* is a nice set of constraints *about* $\neg\phi$, then for every event θ

$$\mathfrak{F}(\mathbf{K}_1 \cup \mathbf{K}_1^*)(\theta | \phi) = \mathfrak{F}(\mathbf{K}_1)(\theta | \phi)$$

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

□

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}_1 contains only information about the *relative* probabilistic beliefs between events in the world of ϕ , while \mathbf{K}_1^* contains only information about beliefs concerning events in the world of $\neg\phi$, then the values which the inference process \mathfrak{F} calculates for probabilities of events conditioned on ϕ should be unaffected by the information in \mathbf{K}_1^* , 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}_1^* since the constraints in \mathbf{K}_1^* may well affect belief in ϕ itself. In essence 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.

As an additional justification for the above principle we may also note the following

Theorem 2.2

The inferences processes **ME**, **CM**[∞], **MD** (minimum distance), together with all **Renyi** inference processes (see e.g. [8]), all satisfy the Locality Principle.

□

The Locality Principle is in essence a generalisation of the Relativisation Principle of Paris [12] and the Homogeneity Axiom of Hawes [8], and the above theorem is very similar to results proved previously, especially to results in [8]. It follows from Theorem 2.2 that if we reject the Locality Principle, then we are in effect forced to reject not just **ME**, but also all the currently most favoured inference processes.

An interesting aspect of the Locality Principle is that the justification given above

appears no less cogent when we attempt to generalise it to a collective situation. If we accept the arguments 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. 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 $\mathbf{A}_1 \dots \mathbf{A}_m$. If for each $i = 1 \dots m$ \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

$$\mathfrak{F}(\mathbf{K}_1 \cup \mathbf{K}_1^*, \dots, \mathbf{K}_m \cup \mathbf{K}_m^*)(\theta | \phi) = \mathfrak{F}(\mathbf{K}_1, \dots, \mathbf{K}_m)(\theta | \phi)$$

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

□

At this point we make a simple observation. In the special 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, it is easy to construct an example to show that if the Locality Principle is to be satisfied then that pooling operator cannot be **LinOp**. On the other hand the pooling operator **LogOp** is perfectly consistent with the Locality Principle, as we shall see in the final section of this paper. 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 the compelling nature of certain arguments in favour of the Locality Principle in the far broader context of a social inference process. Nonetheless, as remarked above, the Locality Principle is violated by the widely used pooling operator **LinOp**, a fact which appears to us to cast serious doubt on the intrinsic plausibility of **LinOp** as a pooling operator.

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 k clones of \mathbf{A}_i , so that we obtain a new college \mathbf{M}^* with km 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. This argument generates the following:

The Proportionality Principle

For any integer $k > 1$

$$\mathfrak{F}(\mathbf{K}_1 \dots \mathbf{K}_1, \mathbf{K}_2 \dots \mathbf{K}_2, \dots, \mathbf{K}_m \dots \mathbf{K}_m) = \mathfrak{F}(\mathbf{K}_1, \mathbf{K}_2, \dots, \mathbf{K}_m)$$

where in the expression on the left there are exactly k copies of each \mathbf{K}_i . □

The proportionality principle looks rather innocent. Nevertheless as we shall see at the end of the next section a slight generalisation 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 social inference process, **SEP**, which satisfies all the principles introduced in the previous section, and which extends both the inference process **ME** and the pooling operator **LogOp**.

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 $\mathbf{K}_1 \dots \mathbf{K}_m$ which respectively represent the beliefs sets of the individuals $\mathbf{A}_1 \dots \mathbf{A}_m$. 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 **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 $\mathbf{A}_1 \dots \mathbf{A}_m$ into a social belief function 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 **M** that his method is optimal.

\mathbf{A}_0 decides that he will choose a social belief function $v = \langle v_1 \dots v_J \rangle$ in such a manner as to minimise the average informational distance between $\langle v_1 \dots v_J \rangle$ and the m belief functions $w^{(i)} = \langle w_1^{(i)} \dots w_J^{(i)} \rangle$ of the members of **M**, where the $w^{(i)}$ are each simultaneously chosen in such a manner as to minimise this

quantity subject to the relevant sets of belief constraints \mathbf{K}_i of each of the members of the college.

Using the standard cross-entropy measure of informational distance this idea amounts to minimising the function

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

subject to all the constraints. In the above the usual convention is observed that $v_i \log \frac{v_i}{w_j^{(i)}}$ takes the value 0 if $v_i = 0$ and the value $+\infty$ if $w_j^{(i)} = 0$ and $v_i \neq 0$.

A little algebraic manipulation establishes that minimising the above expression subject to all the constraints is equivalent to *first* choosing the $w^{(i)}$ subject to the \mathbf{K}_i so as to maximise the function

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

and then, if this maximum value attained is say M , setting

$$v_j = \frac{1}{M} \left(\prod_{i=1}^m w_j^{(i)} \right)^{\frac{1}{m}}$$

for each $j = 1 \dots J$. Notice that the function being maximised above is just a sum of geometric means. Since this function is bounded and continuous and the space over which it is being maximised is by assumption closed, a maximum value M is certainly attained. Moreover it is easy to see that

Lemma 3.1 Given $\mathbf{K}_1 \dots \mathbf{K}_m$ and M defined as above then $0 < M \leq 1$.

Furthermore the value $M = 1$ occurs if and only if for every $j = 1 \dots J$ and for all $i, i' \in \{1 \dots m\}$ $w_j^{(i)} = w_j^{(i')}$. Hence given $\mathbf{K}_1 \dots \mathbf{K}_m$ the following are equivalent:

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

□

Now it is obvious from the above that chairman \mathbf{A}_0 's proposed method of choosing 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 v in this intersection, if adopted as the belief function of each member, will generate the maximum possible value for M of 1 and so will be a possible candidate for a social belief function 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 $w^{(i)}$ or of v .

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

However in order to show that this procedure is well-defined chairman \mathbf{A}_0 needs to prove a number of lemmas.

Definition 3.2

For constraint sets $\mathbf{K}_1 \dots \mathbf{K}_m$ we define

$$M_{\mathbf{K}_1 \dots \mathbf{K}_m} = \text{Max} \left\{ \sum_{j=1}^J \left(\prod_{i=1}^m w_j^{(i)} \right)^{\frac{1}{m}} \mid w^{(i)} \in \mathbf{V}_{\mathbf{K}_i} \text{ for all } i = 1 \dots m \right\}$$

and

$$\Gamma(\mathbf{K}_1 \dots \mathbf{K}_m) = \{ \langle w^{(1)} \dots w^{(m)} \rangle \in \bigotimes_{i=1}^m \mathbf{V}_{\mathbf{K}_i} \mid \sum_{j=1}^J \left(\prod_{i=1}^m w_j^{(i)} \right)^{\frac{1}{m}} = M_{\mathbf{K}_1 \dots \mathbf{K}_m} \}$$

□

By the earlier discussion, each point $\langle w^{(1)} \dots w^{(m)} \rangle$ in $\Gamma(\mathbf{K}_1 \dots \mathbf{K}_m)$ gives rise to a uniquely determined corresponding social belief function v whose j 'th coordinate is given by

$$v_j = \frac{1}{M_{\mathbf{K}_1 \dots \mathbf{K}_m}} \left(\prod_{i=1}^m w_j^{(i)} \right)^{\frac{1}{m}}$$

We will refer to the v thus obtained from $\langle w^{(1)} \dots w^{(m)} \rangle$ as

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

and we let

$$\Delta(\mathbf{K}_1 \dots \mathbf{K}_m) = \{ \mathbf{LogOp}(w^{(1)} \dots w^{(m)}) \mid \langle w^{(1)} \dots w^{(m)} \rangle \in \Gamma(\mathbf{K}_1 \dots \mathbf{K}_m) \}$$

$\Delta(\mathbf{K}_1 \dots \mathbf{K}_m)$ 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.

The following structure theorem for $\Gamma(\mathbf{K}_1 \dots \mathbf{K}_m)$, which depends strongly on the concavity properties the geometric mean function and of sums of such functions, guarantees that Chairman \mathbf{A}_0 's plan is realisable.

Theorem 3.3

For fixed constraint sets $\mathbf{K}_1 \dots \mathbf{K}_m$

(i) For any two points $\langle w^{(1)} \dots w^{(m)} \rangle$ and $\langle \bar{w}^{(1)} \dots \bar{w}^{(m)} \rangle$ in $\Gamma(\mathbf{K}_1 \dots \mathbf{K}_m)$ there exists real numbers $\mu_1 \dots \mu_J \in \mathbb{R}$ such that

$$\bar{w}_j^{(i)} = w_j^{(i)}(1 + \mu_j)$$

for all $i = 1 \dots m$ and $j = 1 \dots J$.

(ii) $\Gamma(\mathbf{K}_1 \dots \mathbf{K}_m)$ is a compact non-empty convex set.

(iii) $\Delta(\mathbf{K}_1 \dots \mathbf{K}_m)$ is a compact non-empty convex set.

(iv) The map

$$\mathbf{LogOp} : \Gamma(\mathbf{K}_1 \dots \mathbf{K}_m) \rightarrow \Delta(\mathbf{K}_1 \dots \mathbf{K}_m)$$

is a continuous bijection.

□

Now since $\Delta(\mathbf{K}_1 \dots \mathbf{K}_m)$ is a closed convex set by 3.3(iii) and since the entropy function

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

is strictly concave over this set, the set contains a unique point v at which the entropy function achieves its maximum value. It follows at once that the following formal definition of **SEP** defines, for every $\mathbf{K}_1 \dots \mathbf{K}_m$ satisfying the conditions of this section, a unique social belief function.

Definition 3.4

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

$$\mathbf{SEP}(\mathbf{K}_1 \dots \mathbf{K}_m) = \mathbf{ME}(\Delta(\mathbf{K}_1 \dots \mathbf{K}_m))$$

□

Theorem 3.5

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

□

It is worth remarking that Theorem 3.3(i) provides a simple sufficient condition for $\Delta(\mathbf{K}_1 \dots \mathbf{K}_m)$ to be a singleton and thus for the application of **ME** in the definition of **SEP** to be redundant:

Theorem 3.6

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

□

An interesting characteristic of **SEP** is that the **ME** 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 satisfies 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 define a member i of the college to be an *ignorant fanatic* if $\mathbf{V}_{\mathbf{K}_i}$ consists of the single point $\langle \frac{1}{j}, \frac{1}{j} \dots \frac{1}{j} \rangle$. Now starting with a college \mathbf{M} of m individuals and constraint sets $\mathbf{K}_1 \dots \mathbf{K}_m$ as before, let us form for any $k \in \mathbb{N}$, a new college \mathbf{M}_k^* of $km + 1$ members, consisting of a single ignorant fanatic together

with k copies of \mathbf{M} . Now one would hope that for a well-behaved social inference process, in applying the social inference process to \mathbf{M}_k^* the effect of the ignorant fanatic would become negligible as $k \rightarrow \infty$, in which case by proportionality in the limit we should get the same answer as we would get by applying the social inference process to \mathbf{M} .

Pleasingly this is exactly what happens for **SEP**. In fact, if we accept the above principle it turns out that we need never invoke the **ME** stage in the definition of **SEP** at all, since by 3.6 the first stage of the definition of **SEP** already guarantees the uniqueness of the social belief function for each \mathbf{M}_k^* , owing to the presence of the ignorant fanatic, while the limit of these social belief functions turns out to be $\mathbf{SEP}(\mathbf{K}_1 \dots \mathbf{K}_m)$. We restate this as the following theorem:

Theorem 3.7

With the notation as above the set Δ_k of solutions for the social belief functions for \mathbf{M}_k^* corresponding to the first stage of applying **SEP** consists, for each k , of a set containing a single probability distribution, say $v^{[k]}$, and furthermore

$$\lim_{k \rightarrow \infty} v^{[k]} = \mathbf{SEP}(\mathbf{K}_1 \dots \mathbf{K}_m)$$

□

Since the left hand side of the above identity does not involve maximum entropy in its definition, we can argue that this shows that the invocation of maximum entropy in the second stage of the original definition of **SEP** is indeed entirely natural.

A suggestive way of interpreting this result is as follows. In order to calculate the social belief function v for \mathbf{M} , chairman \mathbf{A}_0 first minimizes the sum of the cross entropies as in the first stage of the calculation of **SEP**. If this results in a unique solution then that is taken as the social belief function. If the result is not unique then \mathbf{A}_0 adds his own casting constraint set \mathbf{K}_0 as that of an ignorant fanatic⁸ and recalculates, while diluting his own effect as much as possible by imagining that there are k clones of each of the other members of the college, and that $k \rightarrow \infty$. The resulting inference process is just **SEP**.

⁸Of course in this context it may be preferable to replace the designation “ignorant fanatic” by “impartial chair with leadership qualities”.

In conclusion I am grateful to Alena Vencovská for her helpful comments on some of the ideas presented here, and I also wish to thank Hykel Hosni and Franco Montagna for their careful editorial suggestions. The sole responsibility for any errors lies however with the author.

REFERENCES

- [1] R.M.Cooke, *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Environmental Ethics and Science Policy Series, Oxford University Press, New York, 1991.
- [2] 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.
- [3] 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.
- [4] C.Genest, *A conflict between two axioms for combining subjective distributions*. *J. Royal Statistical Society*, 46(3)pp. 403-405, 1984.
- [5] C.Genest and C.G.Wagner, *Further evidence against independence preservation in expert judgement synthesis*. *Aequationes Mathematicae* 32(1) pp. 74-86, 1987.
- [6] 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.
- [7] C.Genest, K.J.McConway, M.J.Schervish, *Characterization of externally Bayesian pooling operators*. *Ann. Statist.* 14(2) pp. 487-501, 1986.

- [8] 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/>
- [9] W.B.Levy and H.Delic, Maximum entropy aggregation of individual opinions. *IEEE Trans. Systems, Man, Cybernetics* 24(4) pp. 606-613, 1994.
- [10] 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.
- [11] J.B.Paris, and Alena Vencovská, A Note on the Inevitability of Maximum Entropy, *International Journal of Approximate Reasoning*, 4, pp. 183-224, 1990.
- [12] J.B.Paris, *The Uncertain Reasoner's Companion - A Mathematical Perspective*, Cambridge University Press, Cambridge, UK, 1994.
- [13] 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.
- [14] J.B.Paris, Common sense and maximum entropy. *Synthese*, vol.117, pp. 75-93, 1999.
- [15] D.M.Pennock and M.P.Wellman, Graphical Models for Groups: Belief Aggregation and Risk Sharing, *Decision Analysis*, Vol. 2, No. 3, September 2005, pp. 148-164, 2005.
- [16] 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.
- [17] Daniel Osherson and Moshe Vardi, Aggregating disparate estimates of chance, *Games and Economic Behavior*, pp. 148-173, July 2006.
- [18] C.Wagner. Aggregating subjective probabilities: Some limitative theorems. *Notre Dame J. Formal Logic* 25(3) pp. 233-240 , 1984.
- [19] 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.