KYBERNETIKA — VOLUME *50* (2014), NUMBER 2, PAGES 175–188

THE IRRELEVANT INFORMATION PRINCIPLE FOR COLLECTIVE PROBABILISTIC REASONING

Martin Adamčík and George Wilmers

Within the framework of discrete probabilistic uncertain reasoning a large literature exists justifying the maximum entropy inference process, **ME**, as being optimal in the context of a single agent whose subjective probabilistic knowledge base is consistent. In particular Paris and Vencovská completely characterised the **ME** inference process by means of an attractive set of axioms which an inference process should satisfy.

More recently the second author extended the Paris–Vencovská axiomatic approach to inference processes in the context of several agents whose subjective probabilistic knowledge bases, while individually consistent, may be collectively inconsistent. In particular he defined a natural multi–agent extension of the inference process **ME** called the social entropy process, **SEP**. However, while **SEP** has been shown to possess many attractive properties, those which are known are almost certainly insufficient to uniquely characterise it. It is therefore of particular interest to study those Paris–Vencovská principles valid for **ME** whose immediate generalisations to the multi–agent case are not satisfied by **SEP**. One of these principles is the Irrelevant Information Principle, a powerful and appealing principle which very few inference processes satisfy even in the single agent context. In this paper we will investigate whether **SEP** can satisfy an interesting modified generalisation of this principle.

Keywords: uncertain reasoning, discrete probability function, social inference process, maximum entropy, Kullback–Leibler, irrelevant information principle

Classification: 03B42, 03B48, 68T37

1. MOTIVATION

In this paper we consider the following fundamental problem of discrete multi-agent probabilistic uncertain reasoning. We are interested in finding a general procedure which, given a finite set of agents, each possessing a subjective probabilistic knowledge base over a finite space of possible events, yields a single probability function or *social probability function* defined over that space of events, which optimally represents the joint knowledge of all the agents, and such that that general procedure satisfies some natural criteria derived from logical or rational considerations.

There are several initial assumptions we want to make. Firstly we assume that the probabilistic beliefs or "knowledge" of each *particular* expert is consistent with the laws

DOI: 10.14736/kyb-2014-2-0175

of probability. Secondly all agents are assumed to have equal status, and the final social probability function should not depend on the order in which the agents' knowledge bases are considered.

We illustrate the motivation behind this idea by a toy two-agent example.

Imagine that two safety experts are evaluating safety in a chemical factory producing nitrogen fertilizers. For simplicity we consider only the ammonia supply which is stored in a tank connected to the rest of the factory by a valve which is controlled by an electric circuit.

The first expert believes that there is a 4% chance that a mechanical problem will cause the valve to fail. The second expert comes up with a different opinion that there is an 8% chance that a mechanical problem will cause the valve to fail. Moreover, the first safety expert thinks that there is a 7% chance that the electric circuit will fail. We suppose that both experts have no other knowledge related to this problem.

Taken together the joint beliefs (knowledge) of the two experts are inconsistent in this case. In practice an individual's knowledge is usually incomplete and offers a lot of uncertainty; the first expert in above example has no knowledge about, for instance, the conditional probability of a mechanical fault on the the valve occurring, given a fault on the electric circuit. The situation becomes more complicated once the second agent is considered, whose knowledge is inconsistent with the knowledge of the first agent. Considering both of the experts' knowledge together we could for example ask the following question:

Question. *How should a rational adjudicator whose only knowledge consists of what is related to him by the two experts above, evaluate the probability that both the valve and the electric circuit will be faulty, based only on the experts' subjective knowledge specified above and without any other assumptions?*

Assuming, as we do in this paper, that each agent's uncertain knowledge can be represented within the framework of probability theory, we can describe the knowledge of each expert by a set of possible probability distributions over four possible mutually exclusive cases: there will be (1) a fault on the valve and a fault on the electric circuit, (2) a fault on the valve and no fault on the electric circuit, (3) no fault on valve and a fault on the electric circuit and (4) no faults on the valve or on the electric circuit. We can denote the corresponding probabilities that (1),(2),(3) and (4) is true by real numbers w_1, w_2, w_3 and w_4 from the interval $[0,1]$ which sum to 1. Based on the knowledge of the first expert $w_1 + w_2 = 0.04$ and $w_1 + w_3 = 0.07$. Any probability function $(x, 0.04 - x, 0.07 - x, x + 0.89)$, where $x \in [0, 0.04]$, is consistent with the knowledge of the first expert. Similarly, the second expert admits any $(x, 0.08 - x, y, 0.92 - y)$ where $x \in [0, 0.08]$ and $y \in [0, 0.92]$. This representation of the knowledge of the experts naturally abstracts from the complex nature of the actual problem. However we are not interested here in the particular manner in which this abstraction from the infinite complexity of a real world problem has been accomplished. Instead we will focus on the following narrower, abstract, but more clearly defined question:

Question. *Given two (or more) sets of probability functions corresponding to the knowledge bases of corresponding experts as in the above example, which single probability function best represents the combined probabilistic knowledge of the experts?*

Naturally, we would like to find a general procedure doing this for any knowledge bases which satisfies some natural principles. We will formalize this idea in a general setting in the next section.

We should emphasize here that our reduction of the experts' knowledge to a theoretical question as above assumes that *all* the knowledge of an individual expert is incorporated in the formal representation of her knowledge base. This last assumption is sometimes referred to as the *Principle of Total Evidence* [2] or the *Watts assumption* (see [11]). In order to avoid confusion, it is clearly essential that this assumption be adhered to in any discussion of the general theoretical characteristics of a mode of probabilistic inference. Of course when applied to the formalisation of any real life problem considered by a *human* agent, the Principle of Total Evidence is never adhered to in practice, as indeed is illustrated by our formalisation of the toy example above. This banal fact of life has historically bedevilled theoretical discussion of probabilistic inference, because it is often extremely hard to give illustrative real world examples of abstract principles of probabilistic inference without a philosophical opponent being tempted to challenge one's reasoning using implicit background information concerning the example which is not included in its formal representation as a knowledge base. This tendency to overlook the Principle of Total Evidence provides an inexhaustible supply of invalid arguments[1].

On the other hand if one assumes that all of each expert's individual knowledge has been included in the formal representation of her knowledge base, it is clearly of considerable interest to formulate and study general principles which tell us under what circumstances part of that knowledge can be considered to be *irrelevant* to the determination of the value which the social probability function should accord to a particular event. It is this idea which forms the central theme of the present paper.

2. FORMALIZATION

Let $L = \{a_1 \dots a_h\}$ be a finite propositional language where $a_1, \dots, a_h$ are propositional variables. In our example $n = 2$, a_1 stands for sentence "there will be a fault on the valve" and a_2 stands for sentence "there will be a fault on the electric circuit". By the disjunctive normal form theorem any L-sentence can by expressed as a disjunction of atomic sentences (atoms) and we will denote a maximal set of logically inequivalent atoms $\{\alpha_1, \dots, \alpha_J\}$, where $J = 2^h$, by $\mathrm{At}(L)$. The atoms of $\mathrm{At}(L)$ are thus mutually exclusive and exhaustive.

A *probability function* $\mathbf{w}$ over L is defined by a function $\mathbf{w} : \mathrm{At}(L) \to [0, 1]$ such that $\sum_{j=1}^{J} \mathbf{w}(\alpha_j) = 1$.

[1] See e.g. Jaynes [6] for an analysis of this phenomenon in relation to certain criticisms of the maximum entropy inference process.

A value of $\mathbf{w}$ on any L-sentence φ may now be defined in the obvious way by setting

$$\mathbf{w}(\varphi) = \sum_{\alpha_j \models \varphi} \mathbf{w}(\alpha_j).$$

Note that formula φ which is not satisfiable, e. g. $a_1 \wedge \neg a_1$, is defined as the disjunction of an empty set of atoms and we set $\mathbf{w}(\varphi) = 0$ in this case.

We will denote the set of all probability functions over L by $\mathbb{D}^L$. For the sake of brevity we will often write w_j instead of $\mathbf{w}(\alpha_j)$, but note that this notation makes sense only for *atomic* sentences α_j. Given a probability function $\mathbf{w} \in \mathbb{D}^L$, a conditional probability is defined by Bayes's formula

$$\mathbf{w}(\varphi|\psi) = \frac{\mathbf{w}(\varphi \wedge \psi)}{\mathbf{w}(\psi)}$$

for any L-sentence φ and any L-sentence ψ such that $\mathbf{w}(\psi) \neq 0$ and is left undefined otherwise.

Now consider two distinct propositional languages $L_1 = \{a_1, \dots, a_{h_1}\}$ and $L_2 = \{b_1, \dots, b_{h_2}\}$. Let $\mathrm{At}(L_1) = \{\alpha_1, \dots, \alpha_J\}$ and $\mathrm{At}(L_2) = \{\beta_1, \dots, \beta_I\}$. Then every atom of the joint language $L_1 \cup L_2$ can be written uniquely (up to logical equivalence) as $\alpha_j \wedge \beta_i$ for precisely one $1 \leq j \leq J$ and precisely one $1 \leq i \leq I$. With only a slight abuse of notation, for an $L_1 \cup L_2$-probability function $\mathbf{r}$ we will often write r_{ji} instead of $\mathbf{r}(\alpha_j \wedge \beta_i)$, in a similar way as for an L_1-probability function $\mathbf{v}$ we write v_j instead of $\mathbf{v}(\alpha_j)$.

Now notice that $\models \alpha_j \leftrightarrow \bigvee_{i=1}^{I} \alpha_j \wedge \beta_i$. Therefore, the marginal probability function whose jth value is given by $\sum_{i=1}^{I} r_{ji}$ is the projection of an $L_1 \cup L_2$-probability function $\mathbf{r}$ to the language L_1. We will denote it by $\mathbf{r}|_{L_1}$. Similarly if Δ is a set of $L_1 \cup L_2$-probability functions, we denote the set $\{\mathbf{v}|_{L_1} : \mathbf{v} \in \Delta\}$ by $\Delta|_{L_1}$. Also if $\mathbf{v}$ is an L_1-probability function and $\mathbf{w}$ is an L_2-probability function then the product function $\mathbf{v} \cdot \mathbf{w}$ defined by $\mathbf{v} \cdot \mathbf{w}(\alpha_j \wedge \beta_i) = v_j w_i$ is an $L_1 \cup L_2$-probability function such that $(\mathbf{v} \cdot \mathbf{w})|_{L_1} = \mathbf{v}$.

A (probabilistic) *knowledge base* $\mathbf{K}$ over L is a set of constraints on probability functions over L such that the set of all probability functions satisfying the constraints in $\mathbf{K}$ forms a nonempty closed convex subset $V_{\mathbf{K}}$ of $\mathbb{D}^L$. $V_{\mathbf{K}}$ may be thought of as the set of possible probability functions of a particular agent which are consistent with her subjective probabilistic knowledge base $\mathbf{K}$. In the sequel we shall loosely identify $\mathbf{K}$ with $V_{\mathbf{K}}$, and may also refer to such a $V_{\mathbf{K}}$ as a knowledge base. Note that the non-emptiness of $V_{\mathbf{K}}$ corresponds to the assumption that $\mathbf{K}$ is consistent, while if $\mathbf{K}$ and $\mathbf{F}$ are knowledge bases then the knowledge base $\mathbf{K} \cup \mathbf{F}$ corresponds to $V_{\mathbf{K}\cup\mathbf{F}} = V_{\mathbf{K}} \cap V_{\mathbf{F}}$. The set of all knowledge bases $V_{\mathbf{K}}$ over L is denoted by CL.

In the toy example, the knowledge of the first expert can be represented by the knowledge base $\mathbf{K}$ which consist of a set of linear constraints on a probability function $\mathbf{w} = (w_1, w_2, w_3, w_4)$ defined over the atomic sentences $a_1 \wedge a_2$, $a_1 \wedge \neg a_2$, $\neg a_1 \wedge a_2$ and $\neg a_1 \wedge \neg a_2$. Then $\mathbf{K} = \{w_1 + w_2 = 0.04,\ w_1 + w_3 = 0.07\}$ and $V_{\mathbf{K}} = \{(x, 0.04 - x, 0.07 - x, x + 0.89) : x \in [0, 0.04]\}$.

Given $\mathbf{K} \in CL_1$ note that the underlying language L_1 is implicitly understood in the notation $V_{\mathbf{K}}$ which should more properly be denoted $V_{\mathbf{K}}^{L_1}$. Thus if $L_1 \subset L$ then $\mathbf{K}$ is also in CL and $V_{\mathbf{K}}^{L} = \{\mathbf{w} \in \mathbb{D}^{L} : \mathbf{w}|_{L_1} \in V_{\mathbf{K}}^{L_1}\}$. For simplicity we shall normally just write $V_{\mathbf{K}}$ when the appropriate language is understood.

We now define the central notion which maps any given sequence of knowledge bases to a single probability function termed the *social probability function* for that sequence. A *social inference process* $\mathcal{S}$ defines for each L and $n \geq 1$ a function

$$\mathcal{S}_L : \underbrace{CL \times \ldots \times CL}_{n} \rightarrow \mathbb{D}^{L}.$$

The number n here intuitively represents the number of distinct agents or distinct sources of information.

The restricted notion $\mathcal{S}$ (or $\mathcal{S}_L$) in the case of a single knowledge base or agent, i.e. when $n = 1$, is simply called an *inference process* and the properties of such inference processes have been extensively studied by Paris, Vencovská and others (see [5, 11, 12, 13, 15]).

As was noted above, a consistent knowledge base $\mathbf{K}$ yields a set $V_{\mathbf{K}}$ of possible probability functions consistent with $\mathbf{K}$. In the case of single agent with knowledge base $\mathbf{K}$ there are several possible procedures to choose a specific probability function from $V_{\mathbf{K}}$. However by the traditional possible worlds modelling or information theoretic arguments whose origins go back to nineteenth century statistical mechanics as in [6] or [12], the maximum entropy inference process **ME** has been justified as being optimal, where $\mathbf{ME}_L(\mathbf{K})$ is defined as that unique probability function $\mathbf{w}$ in $V_{\mathbf{K}}$ which maximizes the Shannon entropy $H(\mathbf{w})$ of $\mathbf{w}$ given by

$$H(\mathbf{w}) = -\sum_{j=1}^{J} w_j \log w_j.$$

H is a strictly concave function and therefore it attains a unique maximum over any nonempty closed convex region $V_{\mathbf{K}}$ of $\mathbb{D}^{L}$.

A quite different justification for **ME** to the traditional ones was described in [15] by Johnson and Shore. Their work was developed by Paris and Vencovská in [13] where they showed that a list of principles based on symmetry and consistency uniquely characterises **ME**. It therefore seems fruitful to look at the axiomatic approach also in the more general context of a social inference process. Accordingly we may ask:

What general principles should a social inference process $\mathcal{S}$ satisfy in order to ensure that for any given knowledge bases, and in the absence of any other information, $\mathcal{S}$ chooses a social probability function according to rational criteria?

We might hope that ultimately such a set of rational principles may determine uniquely a particular social inference process $\mathcal{S}$.

3. LANGUAGE INVARIANCE AND IRRELEVANT INFORMATION

In this section we examine how certain fundamental invariance principles formulated by Paris and Vencovská for an inference process (see [11]) can be extended to the notion of a social inference process.

An obvious question we need to ask regarding social inference processes is whether they depend on the choice of a particular propositional language $L = \{a_1, \ldots, a_h\}$. For fixed $\mathcal{S}$, L, with $\varphi \in SL$ and $\mathbf{K}_1, \ldots, \mathbf{K}_n \in CL$ consider $\mathcal{S}_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)(\varphi)$. It would seem to be irrational to change this value if L is extended to a larger language by adding a set of propositional variables $\{b_1, \ldots, b_k\}$, all distinct from the variables of L, provided that we have not supplied any new knowledge. Following [11] we will formulate this as the following principle:

LI [Language Invariance Principle]. *A social inference process $\mathcal{S}$ satisfies language invariance if whenever L_1 and L_2 are languages with $L_1 \subseteq L_2$ and $\mathbf{K}_1, \ldots, \mathbf{K}_n \in CL_1$, then*

$$\mathcal{S}_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n)(\varphi) = \mathcal{S}_{L_2}(\mathbf{K}_1, \ldots, \mathbf{K}_n)(\varphi)$$

for any L_1-sentence φ.

In the case when $n = 1$ it is well known that several popular inference processes, including **ME**, satisfy **LI** (see [11] or p. 213 of [5] for details).

Following [11] we may also ask a different question in the same vein. What will happen if alongside the new propositional variables, new knowledge concerning these variables is also provided which contains no reference to the old variables. Again, it would seem to be rational that the value of a social inference process on a sentence that is formulated in the original language should not change. This leads us to

IIP [The Irrelevant Information Principle]. *Let $L = L_1 \cup L_2$ where L_1 and L_2 are disjoint propositional languages, and let $\mathbf{K}_1, \ldots, \mathbf{K}_n$ and $\mathbf{F}_1, \ldots, \mathbf{F}_n$ be knowledge bases formulated for the languages L_1 and L_2 respectively. Then for any L_1-sentence φ*

$$\mathcal{S}_L(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n)(\varphi) = \mathcal{S}_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)(\varphi).$$

In the case when $n = 1$ this principle plays a crucial role in the characterisation of **ME** in [13]. Nevertheless, despite its intuitive plausibility this principle is in fact very hard to satisfy; indeed although **ME** satisfies it, almost all other commonly used (single agent) inference processes do not do so (see [11] and [5] for details).

IIP appears even harder for a social inference processes to satisfy. However, in this multi-agent case we might argue that this principle is just too strong. If knowledge provided by agents for the language L_2 is inconsistent then the addition of such new knowledge may provide us with more information on how strongly the agents disagree, which in turn may affect our evaluation of the knowledge concerning L_1. However, if the new knowledge does not change the level of disagreement as is the case when the new knowledge of all the agents is jointly consistent, then the principle of irrelevant information is arguably more justified. Accordingly we formulate:

CIIP [The Consistent Irrelevant Information Principle]. *Let $L = L_1 \cup L_2$ where L_1 and L_2 are disjoint propositional languages. Let $\mathbf{K}_1, \ldots, \mathbf{K}_n$ and $\mathbf{F}_1, \ldots, \mathbf{F}_n$ be* *knowledge bases formulated for the languages L_1 and L_2 respectively, and suppose that* $\mathbf{F}_1, \ldots, \mathbf{F}_n$ *are jointly consistent. Then for any L_1-sentence φ*

$$\mathcal{S}_L(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n)(\varphi) = \mathcal{S}_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)(\varphi).$$

Assuming **LI** this last equation is equivalent to

$$\mathcal{S}_L(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n)(\varphi) = \mathcal{S}_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n)(\varphi).$$

For instance, in the toy example of section 1 the information of both experts about a possible fault on the electric circuit is both consistent and *a priori* irrelevant to the probability that there will be a fault on the valve. Hence if we want to know only the probability that there will be a fault on the valve, then applying the **CIIP** we need consider only the fact that the first expert states that this probability is 4% and the second states that this probability is 8%.

4. THE SOCIAL ENTROPY PROCESS

In this section we define a particular social inference process formulated by the second author in [17] and [18]. The Social Entropy Process **SEP**, is defined by the following two stage process. At the *first stage* we define the set $\Delta_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ as those probability functions $\mathbf{v}$ which globally minimise the sum of Kullback-Leibler divergences (cross-entropies)

$$\sum_{k=1}^{n} \mathrm{CE}(\mathbf{v}, \mathbf{w}^{(k)}) = \sum_{k=1}^{n} \sum_{j=1}^{J} v_j \log \frac{v_j}{w_j^{(k)}} \tag{1}$$

subject only to the conditions that $\mathbf{w}^{(1)} \in V_{\mathbf{K}_1}, \ldots, \mathbf{w}^{(n)} \in V_{\mathbf{K}_n}$, where

$$v_j \log \frac{v_j}{w_j^{(k)}} = \begin{cases} 0 & \text{if } v_j = 0 \text{ and } w_j^{(k)} = 0, \\ \infty & \text{if } v_j \neq 0 \text{ and } w_j^{(k)} = 0. \end{cases}$$

Recall that v_j and $w_j^{(k)}$ stand for $\mathbf{v}(\alpha_j)$ and $\mathbf{w}^{(k)}(\alpha_j)$ respectively, where α_j is an atom and there are J (logically inequivalent) atoms in $\mathrm{At}(L)$.

It is not difficult to see ([18]) that $\Delta_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ is nonempty if *there is some atom α_j such that for no k is it the case that for all* $\mathbf{w} \in V_{\mathbf{K}_k}$ $\mathbf{w}(\alpha_j) = 0$. Under this condition $\Delta_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ is well-defined. From now on we shall consider only n-tuples of knowledge bases $\mathbf{K}_1, \ldots, \mathbf{K}_n$ which satisfy this condition. Note that the definition of a social inference process is not much restricted by such an assumption.

In [18] it is proved that $\Delta_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ is also a closed convex region of $\mathbb{D}^L$ and therefore there is a unique probability function in $\Delta_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ having maximal entropy, and we will denote this function by $\mathbf{ME}_L(\Delta_L(\mathbf{K}_1, \ldots, \mathbf{K}_n))$. Therefore, at the *second stage* of the definition we set $\mathbf{SEP}_L(\mathbf{K}_1, \ldots, \mathbf{K}_n) = \mathbf{ME}_L(\Delta_L(\mathbf{K}_1, \ldots, \mathbf{K}_n))$. It is clear that $\mathbf{SEP}_L$ coincides with $\mathbf{ME}_L$ in the case when $n = 1$ and, it is straightforward to show that **SEP** is language invariant.

The set $\Delta_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ is often a singleton and in that case the second stage is essentially redundant. For instance, this happens whenever $V_{\mathbf{K}_k}$ is a singleton for some

k. The function which maps $\mathbf{K}_1, \ldots, \mathbf{K}_n$ to $\Delta_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ is therefore called the *weak social entropy process* and is denoted by **WSEP**. **WSEP** may naturally be considered as a *merging operator* which merges the evidence of the n agents into a single knowledge (or evidence) base without necessarily picking a unique social belief function[2].

For any $\mathbf{v} \in \Delta_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ there is an n-tuple $\mathbf{w}^{(1)} \in V_{\mathbf{K}_1}, \ldots, \mathbf{w}^{(n)} \in V_{\mathbf{K}_n}$ minimizing $\sum_{k=1}^n \mathrm{CE}(\mathbf{v}, \mathbf{w}^{(k)})$ defined in (1). We will denote the set of all such n-tuples by $\Gamma_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)$.

Lemma 4.1. The following are equivalent:

(i) The probability functions $\mathbf{v}$, $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)}$ minimize (1) subject only to $\mathbf{w}^{(1)} \in V_{\mathbf{K}_1}, \ldots, \mathbf{w}^{(n)} \in V_{\mathbf{K}_n}$.

(ii) $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)}$ maximize $\sum_{j=1}^J \left(\prod_{k=1}^n w_j^{(k)}\right)^{\frac{1}{n}}$, subject only to $\mathbf{w}^{(1)} \in V_{\mathbf{K}_1}, \ldots,$ $\mathbf{w}^{(n)} \in V_{\mathbf{K}_n}$, and $v_j = \dfrac{\left(\prod_{k=1}^n w_j^{(k)}\right)^{\frac{1}{n}}}{\sum_{j=1}^J \left(\prod_{k=1}^n w_j^{(k)}\right)^{\frac{1}{n}}}$ for all $j = 1, \ldots, J$.

For a proof see [18]. We will define the maximal value of $\sum_{j=1}^J \left(\prod_{k=1}^n w_j^{(k)}\right)^{\frac{1}{n}}$ subject to $\mathbf{w}^{(1)} \in V_{\mathbf{K}_1}, \ldots, \mathbf{w}^{(n)} \in V_{\mathbf{K}_n}$ to be $\mathrm{M}_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)$.

The lemma above implies that $\mathbf{SEP}_L$ coincides with the logarithmic (or "normalised geometric mean") pooling operator of decision theory (cf. [3]) in the very special case when each $V_{\mathbf{K}_k}$ defines a single probability function.

In addition to the above pleasing properties, **SEP** satisfies a set of natural principles listed in [17] and [18] similar to those shown to be satisfied by **ME** in [13]. However these are almost certainly not sufficient to characterise **SEP** in the manner in which **ME** was characterised in [13].

Furthermore, although **SEP** is language invariant, it does not satisfy the Irrelevant Information Principle **IIP**. A simple counterexample is provided by the following[3]:

Let $L_1 = \{a_1\}$, $L_2 = \{a_2\}$ and $L = L_1 \cup L_2$. In the following we consider just two agents. Assume that the first agent possesses knowledge $\mathbf{K}_1 = \{\mathbf{w}(a_1) = 0.2\}$, $\mathbf{F}_1 = \{\mathbf{w}(a_2) = 0.2\}$ and the second has knowledge $\mathbf{K}_2 = \{\mathbf{w}(a_1) = 0.3\}$, $\mathbf{F}_2 = \{\mathbf{w}(a_2) = 0.4\}$. Suppose that $\mathbf{w}^{(1)} \in V^{L_1}_{\mathbf{K}_1 \cup \mathbf{F}_1}$ and $\mathbf{w}^{(2)} \in V^{L_2}_{\mathbf{K}_2 \cup \mathbf{F}_2}$. We can identify $\mathrm{M}_L(\mathbf{K}_1 \cup \mathbf{F}_1, \mathbf{K}_2 \cup \mathbf{F}_2)$ defined above by maximizing the following expression for parameters $\mathbf{w}^{(1)}(a_1 \wedge a_2) = x$ and $\mathbf{w}^{(2)}(a_1 \wedge a_2) = y$:

$$\begin{aligned}\mathrm{M}_L(x, y) = \quad & \sqrt{xy} + \sqrt{(0.2 - x)(0.3 - y)} + \sqrt{(0.2 - x)(0.4 - y)} \\ & + \sqrt{(0.6 + x)(0.3 + y)}.\end{aligned}$$

It is the matter of elementary analysis to prove that the above is strictly maximal for $x = 0.12$ and $y = 0.24$. Since $\mathrm{M}_L(0.12, 0.24) = \sqrt{0.08} + \sqrt{0.48}$ it follows that

$$\mathbf{SEP}_L(\mathbf{K}_1 \cup \mathbf{F}_1, \mathbf{K}_2 \cup \mathbf{F}_2)(a_1) = \frac{\sqrt{0.12 \cdot 0.24} + \sqrt{0.08 \cdot 0.06}}{\sqrt{0.08} + \sqrt{0.48}}. \tag{2}$$

[2]The general framework of probabilistic merging operators is investigated in [1].

[3]A counterexample to **IIP** for **SEP** was first found by Soroush Rafiee Rad (private communication, 2010).

On the other hand
$$\mathbf{SEP}_{L_1}(\mathbf{K}_1, \mathbf{K}_2)(a_1) = \frac{\sqrt{0.06}}{\sqrt{0.06} + \sqrt{0.56}}.$$

Later, in corollary 5.4, we will show that **SEP** is language invariant; hence we also have that
$$\mathbf{SEP}_{L}(\mathbf{K}_1, \mathbf{K}_2)(a_1) = \frac{\sqrt{0.06}}{\sqrt{0.06} + \sqrt{0.56}}. \tag{3}$$
Since (2) and (3) are not equal it follows that **SEP** does not satisfy **IIP**.

Since **IIP** in its single agent form played a crucial role in the characterisation of **ME**, this failure could be interpreted as a significant criticism of **SEP**. However, while **IIP** may perhaps be too strong in the multi-agent case, we note that the weaker **CIIP** principle may still be regarded as a natural generalization of the single agent **IIP** since it also reduces to **IIP** for the case $n = 1$.

We shall say that the merging operator **WSEP** satisfies **CIIP** if, whenever $L = L_1 \cup L_2$ where L_1 and L_2 are disjoint propositional languages and $\mathbf{K}_1, \ldots, \mathbf{K}_n$ and $\mathbf{F}_1, \ldots, \mathbf{F}_n$ are knowledge bases formulated for the languages L_1 and L_2 respectively such that $\mathbf{F}_1, \ldots, \mathbf{F}_n$ are jointly consistent, then
$$\mathbf{WSEP}_L(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n) \mid_{L_1} = \mathbf{WSEP}_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n).$$
We prove that **WSEP** satisfies **CIIP** in the following section. However, except in the cases when $\Delta_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ is a singleton, the question as to whether **SEP** also satisfies **CIIP** remains open.

We conclude this section by mentioning some other approaches to the problem of the integration of multi–agent probabilistic evidence which can be found in the literature. In the very special case when each agent specifies a single probability function, a procedure for combining this evidence into a single probability function is called a *pooling operator* by decision theorists. Pooling operators, and in particular the normalised geometric mean pooling operator to which **SEP** reduces in this special case, have been extensively studied from an axiomatic viewpoint (see e. g. [3]). However, except for the work previously cited ([1, 17, 18]), the more general problematic of a social inference process does not appear to have been investigated from an axiomatic standpoint. Nonetheless there exist a number of proposals for what are in essence particular social inference processes, sometimes defined in a somewhat different framework; in this context we should mention [7, 8, 9, 10, 14] and [16]. In particular in [9] and [16] results are proved about iterative convergence procedures based on Kullback–Leibler divergence, one of which yields particular points in $\Delta_L(\mathbf{K}_1, \ldots, \mathbf{K}_n)$. The connection between these algorithms and **SEP** merits further investigation.

5. WSEP SATISFIES CIIP

In what follows we will fix two distinct propositional languages $L_1 = \{a_1, \ldots, a_{h_1}\}$ and $L_2 = \{b_1, \ldots, b_{h_2}\}$. Let $L = L_1 \cup L_2$ and let $\mathrm{At}(L_1) = \{\alpha_1, \ldots, \alpha_J\}$ and $\mathrm{At}(L_2) = \{\beta_1, \ldots, \beta_I\}$.

For $\mathbf{r} \in SL$, to simplify the notation we will often denote $\mathbf{r}|_{L_1}(\alpha_j)$ by $r_{j\cdot}$. We will also denote the conditional probability function $\mathbf{r}(\beta_i|\alpha_j)$ by $r_{i|j}$. It follows that $r_{ji} = r_{j\cdot}r_{i|j}$, i. e. the value r_{ji} can be computed as the product of the projection of $\mathbf{r}$ to L_1 on the L_1-atom α_j and the conditional probability $\mathbf{r}(\beta_i|\alpha_j)$.

Lemma 5.1. Let $w_j^{(k)} \geq 0$ be real numbers for all $1 \leq j \leq J$ and $1 \leq k \leq n$ where $k, j, J, n \in \mathbb{N}$. Then

$$\sum_{j=1}^{J} \left(\prod_{k=1}^{n} w_j^{(k)} \right)^{\frac{1}{n}} \leq \left(\prod_{k=1}^{n} \sum_{j=1}^{J} w_j^{(k)} \right)^{\frac{1}{n}}. \tag{4}$$

Equality holds if and only if either there are real constants $l^{(1)} > 0, \ldots, l^{(n)} > 0$ such that $l^{(1)}(w_1^{(1)}, \ldots, w_J^{(1)}) = l^{(2)}(w_1^{(2)}, \ldots, w_J^{(2)}) = \ldots = l^{(n)}(w_1^{(n)}, \ldots, w_J^{(n)})$ or $\sum_{j=1}^{J} w_j^{(k)} = 0$ for some k.

This lemma is Hölder's inequality, see [4], and it will be very useful in the following proof.

Lemma 5.2. Let $\mathbf{K}_1, \ldots, \mathbf{K}_n \in CL_1$, $\mathbf{F}_1, \ldots, \mathbf{F}_n \in CL_2$ be such that $\mathbf{F}_1, \ldots, \mathbf{F}_n$ are jointly consistent.

(a) If $\mathbf{v} \in \Delta_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ and $\mathbf{t}$ is an L_2-probability function such that $\mathbf{t} \in \bigcap_{i=1}^{n} V_{\mathbf{F}_i}$ then $\mathbf{v} \cdot \mathbf{t} \in \Delta_{L_1 \cup L_2}(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n)$. In particular $\mathbf{F}_1, \ldots, \mathbf{F}_n$ could be empty in which case $\mathbf{t}$ can be arbitrary.

(b) Let $\mathbf{r} \in \Delta_{L_1 \cup L_2}(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n)$. Then $\mathbf{r}|_{L_1} \in \Delta_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n)$. Moreover $\mathrm{M}_{L_1 \cup L_2}(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n) = \mathrm{M}_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n)$.

P r o o f. For a given $\mathbf{v} \in \Delta_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ let $(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(n)}) \in \Gamma_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ be such that $v_j = \frac{\left(\prod_{k=1}^{n} p_j^{(k)}\right)^{\frac{1}{n}}}{\mathrm{M}_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n)}$. Then $\mathrm{M}_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n) = \sum_{j=1}^{J} \left(\prod_{k=1}^{n} p_j^{(k)}\right)^{\frac{1}{n}}$. For a given $\mathbf{r} \in \Delta_{L_1 \cup L_2}(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n)$ let

$$(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)}) \in \Gamma_{L_1 \cup L_2}(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n)$$

be such that $r_{ji} = \frac{\left(\prod_{k=1}^{n} w_{ji}^{(k)}\right)^{\frac{1}{n}}}{\mathrm{M}_{L_1 \cup L_2}(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n)}$.

Let us consider probability functions $\mathbf{w}^{(1)}|_{L_1}, \ldots, \mathbf{w}^{(n)}|_{L_1}$. We will denote $M = \sum_{j=1}^{J} \left(\prod_{k=1}^{n} w_{j\cdot}^{(k)}\right)^{\frac{1}{n}}$. Then $M \leq \mathrm{M}_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ since $\mathrm{M}_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ is maximal. But by the lemma 5.1 also $\mathrm{M}_{L_1 \cup L_2}(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n) \leq M$, hence

$$\mathrm{M}_{L_1 \cup L_2}(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n) \leq \mathrm{M}_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n). \tag{5}$$

(a) Let $\mathbf{t} \in \bigcap_i V_{\mathbf{F}_i}$. We are going to prove that

$$(\mathbf{p}^{(1)} \cdot \mathbf{t}, \ldots, \mathbf{p}^{(n)} \cdot \mathbf{t}) \in \Gamma_{L_1 \cup L_2}(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n). \tag{6}$$

It is easy to see that $\mathbf{p}^{(1)} \cdot \mathbf{t}, \ldots, \mathbf{p}^{(n)} \cdot \mathbf{t}$ satisfy $\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n$ respectively. Moreover,

$$\sum_{j=1,\ldots,J, i=1,\ldots,I} \left(\prod_{k=1}^{n} p_j^{(k)} t_i \right)^{\frac{1}{n}} = \sum_{j=1,\ldots,J, i=1,\ldots,I} \left(\prod_{k=1}^{n} p_j^{(k)} \right)^{\frac{1}{n}} t_i = \mathrm{M}_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n),$$

since $\sum_{i=1}^{I} t_i = 1$. But from (5) we already know that $\mathrm{M}_{L_1 \cup L_2}(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n) \leq \mathrm{M}_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n)$ hence (6) is proved.

(b) By (a) and (5) we have

$$\mathrm{M}_{L_1 \cup L_2}(\mathbf{K}_1 \cup \mathbf{F}_1, \ldots, \mathbf{K}_n \cup \mathbf{F}_n) = M = \mathrm{M}_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n) \tag{7}$$

hence

$$\sum_{j=1,\ldots,J, i=1,\ldots,I} \left(\prod_{k=1}^{n} w_{ji}^{(k)} \right)^{\frac{1}{n}} = \sum_{j=1}^{J} \left(\prod_{k=1}^{n} \sum_{i=1}^{I} w_{ji}^{(k)} \right)^{\frac{1}{n}}.$$

By lemma 5.1 this equality could only occur if for each j there are real constants $l_j^{(1)} > 0, \ldots, l_j^{(n)} > 0$ such that the proportionality

$$l_j^{(1)}(w_{j1}^{(1)}, \ldots, w_{jI}^{(1)}) = l_j^{(2)}(w_{j1}^{(2)}, \ldots, w_{jI}^{(2)}) = \ldots = l_j^{(n)}(w_{j1}^{(n)}, \ldots, w_{jI}^{(n)})$$

holds, or $w_{j\cdot}^{(k)} = \sum_{i=1}^{I} w_{ji}^{(k)} = 0$ holds for some k.

Let us consider the coefficient j to be fixed. If $w_{j\cdot}^{(k)} = 0$ for every k let $\mathbf{q}_{\cdot|j}$ be an arbitrary L_2-probability function with value on ith atom denoted as $q_{i|j}$. Otherwise for $\bar{k}$ such that $w_{j\cdot}^{(\bar{k})} \neq 0$ let us define

$$q_{i|j} = \frac{w_{ji}^{(\bar{k})}}{w_{j\cdot}^{(\bar{k})}}.$$

Obviously,

$$\sum_{i=1}^{I} q_{i|j} = \sum_{i=1}^{I} \frac{w_{ji}^{(\bar{k})}}{\sum_{i=1}^{I} w_{ji}^{(\bar{k})}} = 1$$

and hence $\mathbf{q}_{\cdot|j}$ is a well defined L_2-probability function. Notice that thanks to proportionality the definition does not depend on the choice of $\bar{k}$:

$$\frac{l_j^{(\bar{k})} w_{ji}^{(\bar{k})}}{l_j^{(\bar{k})} \sum_{i=1}^{I} w_{ji}^{(\bar{k})}} = \frac{l_j^{(k)} w_{ji}^{(k)}}{l_j^{(k)} \sum_{i=1}^{I} w_{ji}^{(k)}}.$$

In other words

$$w_{ji}^{(k)} = w_{j\cdot}^{(k)} q_{i|j}. \tag{8}$$

By (7) the projections to L_1 satisfy

$$(\mathbf{w}^{(1)}|_{L_1}, \ldots, \mathbf{w}^{(n)}|_{L_1}) \in \Gamma_{L_1}(\mathbf{K}_1, \ldots, \mathbf{K}_n).$$

Then for L_1-probability function $\mathbf{v}$ defined by $v_j = \frac{\left(\prod_{k=1}^n w_{j\cdot}^{(k)}\right)^{\frac{1}{n}}}{\sum_{j=1}^J \left(\prod_{k=1}^n w_{j\cdot}^{(k)}\right)^{\frac{1}{n}}}$ we have that $\mathbf{v} \in \Delta_{L_1}(\mathbf{K}_1, \dots, \mathbf{K}_n)$.

Moreover,

$$r_{ji} = \frac{\left(\prod_{k=1}^n w_{ji}^{(k)}\right)^{\frac{1}{n}}}{\sum_{j=1}^J \sum_{i=1}^I \left(\prod_{k=1}^n w_{ji}^{(k)}\right)^{\frac{1}{n}}} = \frac{\left(\prod_{k=1}^n w_{j\cdot}^{(k)} q_{i|j}\right)^{\frac{1}{n}}}{\sum_{j=1}^J \sum_{i=1}^I \left(\prod_{k=1}^n w_{j\cdot}^{(k)} q_{i|j}\right)^{\frac{1}{n}}} = v_j q_{i|j},$$

where $r_{j\cdot} = \sum_i v_j q_{i|j} = v_j$ and $r_{i|j} = \frac{r_{ji}}{r_{j\cdot}} = \frac{v_j q_{i|j}}{r_{j\cdot}} = q_{i|j}$ which gives us the required result that $\mathbf{r}|_{L_1} \in \Delta_{L_1}(\mathbf{K}_1, \dots, \mathbf{K}_n)$. □

Theorem 5.3. **WSEP** satisfies **LI** and **CIIP**.

This follows easily from lemma 5.2. Moreover this together with the fact that **ME** is language invariant (see [11]) yields the following:

Corollary 5.4. **SEP** is language invariant.

Theorem 5.5. **SEP** satisfies the **CIIP** in the special case when there is only one probability function in $\Delta_{L_1}(\mathbf{K}_1, \dots, \mathbf{K}_n)$, say $\Delta_{L_1}(\mathbf{K}_1, \dots, \mathbf{K}_n) = \{\mathbf{w}\}$. Note that by theorem 3.8 in [18] this holds whenever at least one of the agents has a knowledge base which fixes a probability function for L_1.

Proof. By lemma 5.2 (b) clearly

$$\mathbf{SEP}_{L_1 \cup L_2}(\mathbf{K}_1 \cup \mathbf{F}_1, \dots, \mathbf{K}_n \cup \mathbf{F}_n)|_{L_1} = \mathbf{r}|_{L_1} = \mathbf{w} = \mathbf{SEP}_{L_1}(\mathbf{K}_1, \dots, \mathbf{K}_n).$$

□

6. CONCLUSION

In this paper we have sought to investigate the Irrelevant Information Principle in the context of multi-agent uncertain reasoning. While this principle plays a crucial role in an axiomatic characterization of **ME** given in [13], we have argued that the most obvious generalization of the Irrelevant Information Principle to the multi–agent context may be too strong. We have proposed an alternative generalization for a social inference process called the Consistent Irrelevant Information Principle (**CIIP**). We have described the promising social inference process **SEP** first formulated in [17] and its weaker counterpart, the merging operator **WSEP**. We have shown that **WSEP** satisfies **CIIP** and that **SEP** satisfies **CIIP** in many cases. The question as to whether **SEP** satisfies **CIIP** in general remains open.

ACKNOWLEDGEMENT

The authors are very grateful to Alena Vencovská for spotting a mistake in the original proof of lemma 5.2. Thanks are also due to the anonymous referees for their helpful comments.

The research leading to these results has received funding from the European Commission's Seventh Framework Programme [FP7/2007-2013] under grant agreement no 238381.

(Received February 25, 2013)

REFERENCES

[1] M. Adamčík and G. M. Wilmers: Probabilistic merging operators. Logique et Analyse (2013), to appear.

[2] R. Carnap: On the application of inductive logic. Philosophy and Phenomenological Research *8* (1947), 133–148.

[3] S. French: Group consensus probability distributions: A critical survey. In: Bayesian Statistics (J. M. Bernardo, M. H. De Groot, D. V. Lindley, and A. F. M. Smith, eds.), Elsevier, North Holland 1985, pp. 183–201.

[4] G. H. Hardy, J. E. Littlewood, and G. Pólya: Inequalities. Cambridge University Press, 1934.

[5] P. Hawes: An Investigation of Properties of Some Inference Processes. Ph.D. Thesis, The University of Manchester, Manchester 2007.

[6] 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 1979.

[7] G. Kern-Isberner and W. Rödder: Belief revision and information fusion on optimum entropy. Internat. J. of Intelligent Systems *19* (2004), 837–857.

[8] J. Kracík: Cooperation Methods in Bayesian Decision Making with Multiple Participants. Ph.D. Thesis, Czech Technical University, Prague 2009.

[9] F. Matúš: On Iterated Averages of I-projections. Universität Bielefeld, Germany 2007.

[10] D. Osherson and M. Vardi: Aggregating disparate estimates of chance. Games and Economic Behavior *56* (2006), 1, 148–173.

[11] J. B. Paris: The Uncertain Reasoner's Companion. Cambridge University Press, Cambridge 1994.

[12] J. B. Paris and A. Vencovská: On the applicability of maximum entropy to inexact reasoning. Internat. J. of Approximate Reasoning *3* (1989), 1–34.

[13] J. B. Paris and A. Vencovská: A note on the inevitability of maximum entropy. Internat. J. of Approximate Reasoning *4* (1990), 183–224.

[14] J. B. Predd, D. N. Osherson, S. R. Kulkarni, and H. V. Poor: Aggregating probabilistic forecasts from incoherent and abstaining experts. Decision Analysis *5* (2008), 4, 177–189.

[15] J. E. Shore and R. W. Johnson: Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. IEEE Trans. Inform. Theory *26* (1980), 1, 26–37.

[16] J. Vomlel: Methods of Probabilistic Knowledge Integration. Ph.D. Thesis, Czech Technical University, Prague 1999.

[17] G. M. Wilmers: The social entropy process: Axiomatising the aggregation of probabilistic beliefs. In: Probability, Uncertainty and Rationality (H. Hosni and F. Montagna, eds.), 10 CRM series, Scuola Normale Superiore, Pisa 2010, pp. 87–104.

[18] G. M. Wilmers: Generalising the Maximum Entropy Inference Process to the Aggregation of Probabilistic Beliefs. available from `http://manchester.academia.edu/GeorgeWilmers/Papers`

Martin Adamčík, Room 2.226, Alan Turing Building, The University of Manchester, Oxford Road, Manchester M13 9PL. UK.
e-mail: maths38@gmail.com

George Wilmers, School of Mathematics, University of Leeds, Leeds LS2 9JT. UK.
e-mail: george.wilmers@gmail.com