Embedding the results of focussed Bayesian fusion into a global context

Jennifer Sander and Michael Heizmann
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB)
Fraunhoferstr. 1, 76131 Karlsruhe, Germany

ABSTRACT
Bayesian statistics offers a well-founded and powerful fusion methodology also for the fusion of heterogeneous information sources. However, except in special cases, the needed posterior distribution is not analytically derivable. As consequence, Bayesian fusion may cause unacceptably high computational and storage costs in practice. Local Bayesian fusion approaches aim at reducing the complexity of the Bayesian fusion methodology significantly. This is done by concentrating the actual Bayesian fusion on the potentially most task relevant parts of the domain of the Properties of Interest. Our research on these approaches is motivated by an analogy to criminal investigations where criminalists pursue clues also only locally. This publication follows previous publications on a special local Bayesian fusion technique called focussed Bayesian fusion. Here, the actual calculation of the posterior distribution gets completely restricted to a suitably chosen local context. By this, the global posterior distribution is not completely determined. Strategies for using the results of a focussed Bayesian analysis appropriately are needed. In this publication, we primarily contrast different ways of embedding the results of focussed Bayesian fusion explicitly into a global context. To obtain a unique global posterior distribution, we analyze the application of the Maximum Entropy Principle that has been shown to be successfully applicable in metrology and in different other areas. To address the special need for making further decisions subsequently to the actual fusion task, we further analyze criteria for decision making under partial information.

Keywords: Information fusion, Bayesian theory, complexity reduction, criminal investigations, local Bayesian fusion approaches, convex optimization, Maximum Entropy Principle, Principle of Minimum Information, decision making under risk, decision making under partial information.

1. INTRODUCTION
Fusing information from different sources is an important capability for all technical systems interacting with their environment. In order to attain the level of situation awareness which is necessary to fulfill a given task, a system usually cannot rely on a single information channel, e.g., a single sensor, but has to consider multiple information sources. The pieces of information to be fused may relate to the same physical property in the same context (also referred to as homogeneous information) or may only be combinable at a higher level of abstraction (so-called heterogeneous information), e.g., in the context of objects. In addition, prior knowledge has to be incorporated. It may refer to existing experience (expert knowledge), but also to natural laws. Examples for the former case are street maps (if, e.g., vehicles have to be detected or tracked) or constraints (e.g., the empirical knowledge that all vehicles have certain maximum speeds). An example for the latter case is the connection between different components of the state vector of a vehicle representing its dynamics (e.g., position and speed).

The need to fuse different pieces of information is caused by several demands: information fusion may be necessary due to restrictions of the sensors used (e.g., discretization and quantization issues, disturbances and noise). In this context, information fusion leads to higher accuracy and reliability. Combining several information sources may be useful to better characterize the object of interest, e.g., in a classification task. Furthermore, fusing information may speed up the information gathering process and lead to a more cost-effective information acquisition.

Further author information:
Jennifer Sander: E-mail: jennifer.sander@iosb.fraunhofer.de, Telephone: +49 721 6091-546
Michael Heizmann: E-mail: michael.heizmann@iosb.fraunhofer.de, Telephone: +49 721 6091-329
The basic principle of fusing information in order to obtain a better understanding of the world is similar to the way human beings interact with their environment. If one of the information sources is missing (one of the senses or prior knowledge represented in the human’s memory), the cognitive skills of the respective person are severely reduced.

Among different technical approaches to fuse information, the Bayesian approach is outstandingly attractive. Bayesian fusion is totally based on probability distributions acting as information carrier for all information contributions (sensory information and prior knowledge). In comparison to other approaches (e.g., Dempster Shafer theory or Neural Networks), probability distributions are easier understandable.

However, Bayesian fusion suffers from the high costs necessary to obtain the fusion result. In consequence, it is often not possible to apply the pure Bayesian calculus to real world fusion problems. Broadly speaking, the problem arises from the need to determine and to combine probability distributions over the whole domains of the Properties of Interest and the information sources, respectively. In consequence, approaches to reduce the computational and storage costs of Bayesian fusion are necessary. In this contribution, we focus on local Bayesian fusion. The main idea of local Bayesian fusion is that the calculus may be concentrated on certain parts of the domain of the Properties of Interest in which the fusion result is expected. Other parts of the domain of the Properties of Interest are less relevant for the fusion result and are thus treated with less thoroughness. However, due to this negligence of information contributions outside the focus areas, the resulting posterior distribution is no longer identical to the result that would be obtained by applying the pure Bayesian approach. In consequence, the question arises how the results of a local Bayesian fusion can be adequately further used.

This paper is organized as follows: first, we give a short introduction to Bayesian fusion and the related costs in Section 2. Then, Section 3 concentrates on local Bayesian fusion approaches, in particular on a special local Bayesian fusion technique termed focussed Bayesian fusion. In Section 4, the question how a unique and complete posterior distribution can be obtained using local Bayesian approaches is addressed and the explicit application of the Maximum Entropy Principle, an established technique usable to this aim, is shown. Finally, Section 5 deals with the aspect of how the obtained posterior distribution can be further used directly in order to derive subsequent decisions.

This publication follows previous publications of the authors on local Bayesian fusion approaches. In particular, Section 4 takes up respective important results of a technical report. In addition, the primary concepts behind the Maximum Entropy Principle and its concrete working mechanism are described and analyzed more precisely in this section. Also, following a previous publication concentrating purely on the needed mathematical derivations, Section 5 presents the concepts and strategies underlying decision making on basis of a focussed Bayesian model in more detail.

2. BAYESIAN FUSION

2.1 Fundamentals
Let \( z = (z_1, \ldots, z_N) \in Z, N \in \mathbb{N} \), denote the Properties of Interest that describe the aspects (e.g., of a scene, a process, or a situation) that are of interest in a certain task. Their acquisition via an information source generally comes along with an information reduction leading to uncertainty with regard their “true” value. By the fusion of several information sources, this uncertainty is often reducible to an acceptable level. In the following, assume that \( S \in \mathbb{N} \) information sources delivering information with regard to the vector \( z \) are to be fused. Therefore, let \( d_s \in D_s, s \in \{1, \ldots, S\} \), denote the contribution of information source number \( s \) such that the vector \( d = (d_1, \ldots, d_S) \in D = D_1 \times \ldots D_S \) comprises all source specific information.

Bayesian fusion follows the principles of inductive inference by using the information embodied by \( d \) to draw conclusions about the underlying cause \( z \). Thereby, uncertainties are modeled and propagated probabilistically. In essence, the necessary probabilistic calculus results directly from the definition of absolute probability done by the Kolmogorov axioms and the common definition of conditional probability. It is important to note that these mathematical definitions of probability do not stipulate what probability means, i.e., how probability statements are to be interpreted. In the Bayesian approach, probability is interpreted as Degree of Belief, i.e., as quantification of the degree of knowledge one has in the occurrence of considered events. This wide reaching interpretation of probability is an important reason why the Bayesian approach is able to provide a general
concept for modeling uncertainty and for propagating it for example through the different levels of the JDL fusion model.\textsuperscript{5}

Also, it enables the Bayesian approach to treat all quantities being involved in a fusion task as random quantities. This makes it possible to define the underlying Bayesian model via introducing the joint probability distribution $p(z, d)$ on the product space $Z \times D$. The joint distribution factorizes into $p(z|d)p(d)$ and $p(d|z)p(z)$. These two products as well as their identity are the essential elements of the Bayesian approach.

Being based on the interpretation of probability as Degree of Belief, the Bayesian approach expresses the information being available about each considered quantity via the respective conditional probability distribution. Especially, in the strict sense, the result of a Bayesian fusion task is the posterior distribution $p(z|d)$. The posterior distribution expresses all the information being available about $z$ in a concrete fusion task in a probabilistic manner. On basis of $p(z|d)$, particular conclusions are then possible subsequently, e.g., via the calculation of point or interval estimations regarding the “true” values of the Properties of Interest or via incorporation of $p(z|d)$ into a decision theoretic framework for choosing among different alternatives.

The posterior distribution is calculated via the application of the Bayesian theorem which can be shortly expressed as

$$p(z|d) \propto p(d|z)p(z). \quad (1)$$

The Bayesian approach handles prior knowledge being available independently of the information delivered by the $S$ information sources in an explicit manner. The prior knowledge gets incorporated into the Bayesian model via the prior distribution $p(z)$. The other quantity of the right side of Equation (1), $p(d|z)$, captures the causal relation between the information contributions provided by the $S$ information sources and the Properties of Interest. In a concrete fusion task, $d$ is fixed and $z$ is variable. Under this premise, $p(d|z)$ is called Likelihood function. Note that the Likelihood function is not normalized, i.e., it does not integrate to one\textsuperscript{*} (with respect to $z$).

If the components of $d$, i.e., the individual contributions $d_1, \ldots, d_S$ of the information sources are conditionally independent given $z$, $p(d|z)$ can be obtained as the product of component specific Likelihood functions $p(d_1|z), \ldots, p(d_S|z)$. As consequence, the transformation of the source specific information into the needed probabilistic representation corresponding to $p(d|z)$ can be done individually for each of the information sources.

Assuming conditional independence is often justifiable especially in the case that heterogeneous information sources using different physical principles for information acquisition are to be fused.

### 2.2 Costs

Bayesian theory delivers a mathematically founded fusion methodology possessing many advantages in comparison to alternative fusion methods. Particularly worthy of mention are its consistency and its intuitive understandability also for not specially trained persons. However, in many real world tasks, the application of Bayesian fusion is prohibitive due to its high computational and storage costs. The reason therefore is that, except in special cases, the posterior distribution $p(z|d)$ is not analytically derivable. If $Z = Z_1 \times \ldots \times Z_N$ and if each subspace $Z_n$ possesses (possibly after discretization) the cardinality $|Z_n|$, the costs for obtaining the posterior distribution $p(z|d)$ are

$$O(|Z|) = O\left(\prod_{n=1}^{N} |Z_n|\right) = O(\zeta^N) \quad (2)$$

with $\zeta = \sqrt[|Z|]{\prod_{n=1}^{N} |Z_n|}$ denoting the geometric mean value of $|Z_n|$.

In Bayesian theory, the most prominent traditional way to address unacceptable high costs is to avoid them by applying the concept of conjugate priors. Thereby, the freedom one has with regard to building the underlying

\textsuperscript{*}In practice, the components of $z$ as well as the components of $d$ may be partially discrete and partially continuous quantities. It is a clear advantage of the Bayesian approach that it treats both kinds of quantities simultaneously and even in an equal manner. Consequently, in this publication, we will adopt a unified integral notation for both, continuous as well as discrete quantities – unless otherwise stated explicitly. In the case of a discrete quantity, integration is then to be interpreted as summation.
Bayesian model gets restricted significantly. Loosely speaking, one restricts oneself on Bayesian models in that prior and posterior distribution belong to the same distribution class. Additionally, this distribution class must be parametrized by a moderate number of parameters. By this, the derivation of the posterior distribution becomes analytically possible – it reduces itself to an update of these parameters.

Generally, the traditional ways are not sufficient to address the problems corresponding to the potentially high costs of Bayesian fusion. This fact is reinforced especially by the observation that the practical application of Bayesian methods has taken a major boost since modern stochastic simulation methods, especially Markov Chain Monte Carlo (MCMC) methods, have become available and are applied in the Bayesian context. Here, theoretical quantities (like the posterior distribution itself, its moments etc.) are approximated by its empirical counterparts which are obtained by certain sampling methods. Despite the fact that stochastic simulation methods are very successfully applied in a lot of real world problems, it has also become clear that they also do not provide a universal solution in the context of Bayesian fusion.\(^6\)

Until today, there exists no universal solution for keeping the costs of Bayesian fusion on an acceptable level and there are still practically relevant problems so that the existing approaches are not suitable. As consequence, research investigating new approaches for controlling the costs of Bayesian fusion surely possesses a lot of potential.

3. LOCAL BAYESIAN FUSION APPROACHES

3.1 Local consideration of fusion tasks

The basic idea of local Bayesian fusion is that in order to obtain a fusion result, it is generally not necessary to consider the entire domains of the Properties of Interest and the information sources, respectively. Since the fusion result is usually a statement on a limited area (e.g., the detection of an object at a certain position), it should be sufficient to concentrate the fusion effort to such areas where the fusion result can be expected. Such suspicious areas can be identified by means of the information sources themselves, e.g., using actual sensor data. Of course, it is nevertheless necessary to make sure that the fusion result lies in fact within the chosen local area, which means in the context of probabilities that the probability that the result lies outside of the considered area must also be assessed in a suitable manner.

This idea has a similarity with the procedure of criminal investigations. Here, the criminal investigator follows clues, starting from the facts found, e.g., at the crime scene. These clues establish the first local part of the world that is considered to be the focus of investigations (i.e., the local part of the domain of the Properties of Interest). In the path of the investigations, the investigator tries to confirm or disprove the reasoning developed up to then by using all sensory information sources he finds suitable for the given context and by incorporating pertaining prior knowledge. In this respect, context related information means only information that refers to the considered local part of the world, e.g., suspects and their environment. No criminal investigator would take account of evaluating all conceivable information from all over the world – the task of criminal investigation only becomes manageable by means of the restricted focus of the investigations. However, for the validity of the investigation result, it is crucial to know how reliable the outcome is. This means that there must be a statement on how likely the investigation result is correct (i.e., that the real issue was in fact within the considered local area of investigations), and the certainty that the result is correct must be high.

This approach of local fusion has also an analogy in the human cognition: when a human being is interested in an actual situation, he will guide his attention such that his senses preferentially cover the areas of his environment where something interesting may happen. A good example is the participation of a human in road traffic, e.g., as a pedestrian. A human being will then intuitively direct his senses towards potential hazards, e.g., directions where vehicles may come from. When he spots something that draws his attention (when he hears a suspicious noise, for example), he will automatically gaze at the area where he expects something to happen. Information from other directions may be neglected in this situation, since he concentrates all his senses to this particular event. However, it is essential for surviving the illustrated traffic example that the human being has a high enough certainty that the event is really where he expects it to be, and to change his focus of interest if the information is not certain enough. That means that, in order to survive, a human being will not totally restrict all his senses on the chosen local area.
3.2 Fundamental mathematical considerations

The basic idea behind local Bayesian fusion approaches is to concentrate the given fusion task on a local context $U \subset Z$, i.e., an adequately chosen subset of the range of values of the Properties of Interest. $U$ must contain at least these values of the Properties of Interest that have a higher potential to be the “true” value of $z$ than other values have. To point out the analogy to criminal investigations more concretely, $U$ must contain at least these values of $z$ which are indicated by clues.

An important challenge regarding the research on local Bayesian fusion was to translate this consideration into a mathematical notion. It has been shown by the authors that, from a mathematical point of view, $U$ can be specified adequately if it contains at least these values of $z$ for which the standardized Likelihood function is larger than a (suitably chosen) threshold\(^1\), i.e.,

$$\frac{p(d|z)}{\max_{z \in Z} p(d|z)} > \gamma, \gamma \in (0,1).$$

(3)

Alternatively, especially for the case that the information contributions $d_s$ are conditionally independent given $z$, it is highly reasonable to choose $U$ such that it contains at least these values of $z$ for which at least one of the source specific standardized Likelihood functions is larger than a (suitably chosen) threshold\(^1\), i.e.,

$$\frac{p(d_s|z)}{\max_{z \in Z} p(d_s|z)} > \delta, \delta \in (0,1).$$

(4)

The surely most important rationale for this proceeding results from the concept of misleading evidence.\(^8\) From this, a quantitative probabilistic error bound regarding the validity of the corresponding local Bayesian model has been derived.\(^7,9\) In addition to this, also other quality indicators regarding the validity of the corresponding local Bayesian model have been investigated by the authors.\(^10\)

There are different answers to the question how to model the concentration of the given fusion task on the local context $U$ mathematically.\(^9\) One possible way is to perform the fusion with regard to $U$ in detail and to perform it with regard to the rest of $Z$, i.e., with regard to $Z \setminus U$, only in a coarsened way. This means, e.g., that for each $z \in U$ the posterior distribution $p(z|d)$ and in addition the total posterior probability $P(Z \setminus U|d) = \int_{Z \setminus U} p(z|d)dz$ of the complement of the local context $U$ are calculated. The other way is to restrict the actual fusion task completely on the local context $U$ which means ignoring all values contained in the set $Z \setminus U$ completely. From a mathematical point of view, this means to condition the Bayesian model on the assumption that $z \in U$ holds. As previous research has shown, except in special cases, the second way is the more practicable way. The present publication concentrates on the respective second local Bayesian fusion technique which we termed focussed Bayesian fusion. Of course, the fact that, mathematically, $Z \setminus U$ is ignored completely in the respective focussed Bayesian model implies the need for other means to consider $Z \setminus U$ when interpreting and analyzing the results of focussed Bayesian fusion (compare Section 3.1).

Focussed Bayesian fusion distinguishes itself by a practically uncomplicated fusion scheme.\(^10\) The corresponding posterior distribution $p(z|d, U)$ can be obtained as follows:

$$p(z|d, U) \propto p(d|z)p(z) \propto p(d|z)p(z|U).$$

(5)

This means that it is sufficient to represent the prior knowledge probabilistically only with regard to $U$. Also, the Likelihood function of the respective focussed Bayesian model corresponds simply to an extract of the global (i.e., the original) Bayesian model. This means that the causal relationship between the information contributions provided by the $S$ information sources and the Properties of Interest can get captured as usual but it suffices to capture it for these values of the Properties of Interest that are contained in the local context.

---

\(^1\)In essence, this corresponds to thresholding the Likelihood function $p(d|z)$, here.

\(^2\)By the standardization, the individual Likelihood functions get advantageously aligned, here.\(^7\)
3.3 Connection between the global and the focussed Bayesian model

From a global view, the conditioning on \( U \) means that all events \( E \subseteq Z \setminus U \) get assigned posterior probability (as well as prior probability) zero. This means that the probability mass lying on \( Z \setminus U \) in the global Bayesian model gets shifted on \( U \) in the focussed Bayesian model. By this, the posterior probability (and the prior probability, respectively) of events \( E \subseteq U \) gets increased according to \( \frac{1}{P(U|d)} \) (and according to \( \frac{1}{P(U)} \), respectively). More concretely, it holds for each event \( E \subseteq Z \)

\[
P(E|d,U) = \int_E p(z|d,U)dz \tag{6}
\]

with

\[
p(z|d,U) = \begin{cases}
p(z|d), & \text{if } z \in U, \\
0, & \text{if } z \notin U.
\end{cases} \tag{7}
\]

In particular, it holds for events \( E \subseteq U \)

\[
P(E|d) = P(E|d,U)P(U|d) \leq P(E|d,U) \tag{8}
\]

and for events \( E \subseteq Z \setminus U \)

\[
P(E|d) \leq 1 - P(U|d) . \tag{9}
\]

It is important to note that it is not possible that the (unknown) global posterior probabilities (as well as the prior probabilities) of events being contained in the local context \( U \) vary arbitrarily. According to Equation (6) and Equation (7), it holds for events \( E, F \subseteq U \) that

\[
\frac{P(E|d,U)}{P(F|d,U)} = \frac{P(E|d)}{P(F|d)} =: r(E,F) . \tag{11}
\]

Let the event \( F \subseteq U \) be arbitrary but fixed. If we assume that \( P(F|d) \) takes a certain value, \( P(E|d) \) is uniquely determined for each other event \( E \subseteq U \) because, according to Equation (11), we have

\[
P(E|d) = P(F|d)r(E,F) . \tag{12}
\]

Thereby, \( r(E,F) \) is computable within the focussed Bayesian model.

The results presented so far in this section may be sufficient for deriving all conclusions needed in a certain fusion task completely on basis of the focussed Bayesian model\(^9\). It must be stressed that, especially in this case, good quality indicators regarding the validity of the focussed Bayesian model (compare Section 3.2) are essential to minimize the danger of erroneous results.\(^9\)

In this publication, we primarily derive and contrast different ways of embedding the results of focussed Bayesian fusion more explicitly into a global context. To this aim, the knowledge how the translation from the global Bayesian model to the focussed Bayesian model has been done will be used advantageously. More concretely, the knowledge that the local context \( U \) has been chosen according to Equation (3) delivers a lower bound for its global posterior probability.\(^1\) It holds

\[
1 \geq P(U|d) \geq \frac{\int_U p(d|z)p(z|U)dz}{\int_U p(d|z)p(z|U)dz + (\frac{1}{P(U)} - 1)} \tilde{\gamma} =: \beta, \quad \tilde{\gamma} := \gamma \cdot \max_z p(d|z) . \tag{13}
\]

If \( U \) has been chosen according to Equation (4), a similar lower bound for \( P(U|d) \) results.\(^2\) Note that the choice of \( U \) may prove itself to be critical in cases in that a rather low value for \( \beta \) is obtained. Obviously, such a situation may occur as consequence of a rather imprecise estimation of \( P(U|d) \) or of the fact that \( P(U|d) \) is truly rather low.
Inserting Equation (13) in Equation (8), we obtain a lower bound for the global posterior probability of events $E \subseteq U$. Combining this result with the upper bounds according to Equation (9) and Equation (10), the following probability interval scheme for global posterior probabilities results:

$$P(E|d) \in \begin{cases} [\beta \cdot P(E|d, U), P(E|d, U)], & \text{if } E \subseteq U, \\ [0, 1 - \beta], & \text{if } E \subseteq Z \setminus U. \end{cases} \quad (14)$$

Equation (12) and Equation (14) (or, alternatively, Equation (7) and Equation (13)) summarize the knowledge one has with regard to the global posterior probabilities when performing focused Bayesian fusion. In the following two sections, we will derive and contrast different ways of explicitly embedding this knowledge into the respective global context.

## 4. DERIVING A UNIQUE GLOBAL POSTERIOR DISTRIBUTION

As we have seen in Section 3.3, focused Bayesian fusion delivers partial information with regard to the global posterior distribution $p(z|d)$. Concretely, we know that $p(z|d)$ must be such that Equation (12) and Equation (14) (or, alternatively, Equation (7) and Equation (13)) are fulfilled. This information is called partial, as it does not uniquely determine $p(z|d)$, i.e., different probability distributions will be consistent with these constraints.

The global posterior distribution $p(z|d)$ probabilistically represents the uncertainty remaining with regard to the “true” value of $z$ after the fusion has been done. In contrast, the intervals for the global posterior probabilities according to Equation (14) represent (together with Equation (12)) the uncertainty resulting additionally due to the transition from the global to the focused Bayesian model in an explicit non-probabilistic manner. It is common sense in the Bayesian community that – at least provided that such a distinction of uncertainties is not used effectively – all facts and corresponding uncertainties are represented adequately by probability in sense of the Degree of Belief interpretation. According to this, on basis of the available information, a unique global posterior distribution should be derived.

Bayesian theory provides several techniques for adequately incorporating certain information into a probabilistic representation in sense of the Degree of Belief interpretation. In this publication, we will concentrate ourselves on the Maximum Entropy Principle, an established technique being successfully applied in metrology as well as in multiple other research disciplines.

In Section 4.2, we will apply the Maximum Entropy Principle to derive a unique global posterior distribution $p(z|d)$ on basis of the partial knowledge being available with regard to this quantity. Before that, in Section 4.1, we will introduce the primary concepts behind this technique for incorporating certain information into a probabilistic representation. From this introduction, the reader will see that the meaningfulness of the Maximum Entropy Principle directly lies on hand for the discrete case. It will also become clear that its extension to the continuous case is somewhat more complicated as this extension needs more theoretical justification. Additionally, it involves some arbitrariness to be overcome. However, it is emphasized that, despite the difficulties, the Maximum Entropy Principle provides a very powerful technique for obtaining a unique probabilistic representation of given information such that its subsequent application in Section 4.2 is surely worth to be taken into account.

### 4.1 Primary concepts behind the Maximum Entropy Principle

If $z$ is a discrete random quantity with values in $Z$ and distributed according to $q(z)$, the corresponding (Shannon) entropy is given by

$$H[q(z)] = -\sum_z q(z) \log q(z). \quad (15)$$

$H[q(z)]$ is a quantitative measure for the amount of uncertainty or, in other words, an inverse measure for the amount of information contained in $q(z)$. This statement can be derived directly from axioms for a meaningful measure for the amount of uncertainty.\(^1\)

\(^{1}\)Contrary to the rest of this publication, an explicit distinction between discrete and continuous quantities is needed in the first part of this section.
Usually, different probability distributions are consistent with given information. The task is then to make an adequate choice among the resulting set of possible probability distributions. Especially in the context of information fusion, one usually intends that the chosen probability distribution is objective in the sense that it represents a kind of common understanding regarding the state of information and not the subjective view of an individual person. To reach this aim, one must choose that probability distribution which is on the one hand consistent with the given information and which incorporates no additional information ("artifacts") on the other hand. Connecting this observation with the notion of entropy, the task is choosing from all probability distributions being consistent with the given information that probability distribution possessing maximal entropy.

In practice, the Maximum Entropy Principle is also often applied successfully when continuous quantities are involved. In the continuous case, the (Boltzmannn) entropy of a (continuous) random quantity \( z \) with values in \( Z \) and distribution \( q(z) \) is given by

\[
H[q(z)] = -\int_Z q(z) \log q(z) \, dz .
\]  

(16)

It must be mentioned, here, that the interpretation of the continuous entropy as quantitative measure for the amount of uncertainty raises conceptual difficulties. Purely formally, Equation (16) results from Equation (15) simply by replacing summation by integration. However, a closer examination shows that Equation (16) does not result as limit of Equation (15) if an interval \( X \) gets discretized increasingly finer.\(^{18, 20}\) Also, unlike the discrete entropy, the continuous entropy may be negative and, unlike its discrete counterpart, the continuous entropy is not invariant under coordinate transformations.\(^{13, 21}\)

A common way to explicitly address these difficulties is to define not an absolute but a relative quantitative uncertainty measure. To this aim, the Kullback Leibler divergence\(^{22}\) (also called relative entropy) that does not possess such critical characteristics can be used for both, the discrete and the continuous case.

If \( q(z) \) and \( r(z) \) are probability distributions on \( Z \), the Kullback Leibler divergence between \( q(z) \) and \( r(z) \) is given by\(^{4}\)

\[
K[q(z), r(z)] = \int_Z q(z) \log \frac{q(z)}{r(z)} \, dz .
\]  

(17)

Suppose that \( r(z) \) is a probabilistic representation of certain information. Assume further that new information, i.e., information becoming available additionally, leads to the replacement of the probabilistic representation \( r(z) \) by the probabilistic representation \( q(z) \). Then \( K[q(z), r(z)] \) serves as a quantitative measure for the additional amount of information being contained in \( q(z) \) (relative to \( r(z) \)).

Now assume that a probabilistic representation \( r(z) \) of certain information \( I_1 \) as well as new information \( I_2 \) are given and that the task is to switch from \( r(z) \) to a probability distribution \( q(z) \) that incorporates additionally also \( I_2 \) in an objective manner. In this case, one must chose that probability distribution which is on the one hand consistent with the total information being available (i.e., \( I_1 \) and \( I_2 \)) and which incorporates no other information on the other hand. These considerations directly lead to the Principle of Minimum Information\(^{23}\) which states that one must chose \( q(z) \) such that it is consistent with the additional information \( I_2 \) while minimizing \( K[q(z), r(z)] \).

There is a very close link between the Principle of Minimum Information and the Bayesian approach. This gets apparent especially from the fact the Bayesian updating rule (according to the Bayesian theorem) results as special case of the Principle of Minimum Information. If \( I_2 \) corresponds to the event that an observed quantity surely adopts a certain value \( d \), the Principle of Minimum Information delivers the result that \( r(z) \) must be updated (according to the Bayesian theorem) to \( r(z|d) \).

Applying the Principle of Minimum Information in cases in that absolutely no information is available until \( I_2 \) arrives leads to the surely most critical point in Bayesian theory, i.e., the choice of a probability distribution representing the fact that we know absolutely nothing. Here, \( r(z) \) must be chosen as being non-informative. Bayesian theory provides different kinds of such non-informative distributions and, of course, the concrete choice

\(^4\) At this point, we come back to the convention that, generally in this publication, a unified integral notation is used for both, discrete and continuous quantities.
influences the value of $K[q(z), r(z)]$. Loosely speaking, the requirement that $r(z)$ shall be non-informative means that this probability distribution shall not favor any value contained in $Z$. Often, it is a good strategy to chose $r(z)$ as being constant, i.e., to set $r(z) \propto 1$, here. This statement is also valid for the case when the resulting $r(z)$ will be an improper probability distribution\(^\text{1}\), i.e., does not integrate to one.\(^\text{24,25}\)

It is not hard to see that, for the case that $r(z)$ is constant, the Principle of Minimum Information directly leads to the Maximum Entropy Principle. In this sense, the Principle of Minimum Information provides a theoretical justification for the use of the Maximum Entropy Principle for both, the discrete and the continuous case. In summary, it is meaningful to interpret the (Shannon and the Boltzmann, respectively) entropy as quantitative measure for the amount of uncertainty (or conversely, as inverse measure for the amount of information) that remains (or conversely, that arises) if we move from an initial non-informative distribution being constant to a new probabilistic representation.\(^\text{18}\)

### 4.2 Application of the Maximum Entropy Principle at focussed Bayesian fusion

Let $\Pi$ denote the set of all probability distributions (denoted by $p(z|d)$) being consistent with Equation (12) and Equation (14) (or, alternatively, with Equation (7) and Equation (13)). Mathematically, for the corresponding Maximum Entropy distribution $p_{\text{ME}}(z|d)$, it holds:

$$p_{\text{ME}}(z|d) = \underset{p(z|d) \in \Pi}{\operatorname{arg\,max}} \; H[p(z|d)] \quad \text{with} \quad H[p(z|d)] = - \int_z p(z|d) \log p(z|d)dz.$$  \hspace{1cm} (18)

This section aims at deriving an explicit formula for $p_{\text{ME}}(z|d)$. For this purpose, we need the following lemma.

**Lemma 1.** Let $X$ denote the set $\{U, Z \setminus U\}$ which is a coarsened version of the range of values of the Properties of Interest. Let $p_{\text{C}}(z|d)$ denote a probability distribution on $X$ that assigns the local context $U$ probability $P(U|d)$. Then, the following connection holds:

$$H[p(z|d)] = H[p_{\text{C}}(x|d)] + P(U|d)H[p(z|d,U)] + P(Z \setminus U|d)H[p(z|d,Z \setminus U)].$$  \hspace{1cm} (19)

**Proof of Lemma 1.** For each set $V \in \{U, Z \setminus U\}$ and for each value $z \in V$, it holds that $p(z|d) = P(V|d)p(z|d,V)$ (compare basically Equation (7)). Inserting this connection into the definition of $H[p(z|d)]$ after splitting the respective integral in two parts and subsequently rearranging the involved terms directly delivers Equation (19):

$$H[p(z|d)] = - \int_z p(z|d) \log p(z|d)dz$$  \hspace{1cm} (20)

$$= - \int_U p(z|d) \log p(z|d)dz - \int_{Z \setminus U} p(z|d) \log p(z|d)dz$$  \hspace{1cm} (21)

$$= - \int_U P(U|d)p(z|d,U) \log (P(U|d)p(z|d,U))dz$$

$$- \int_{Z \setminus U} P(Z \setminus U|d)p(z|d,Z \setminus U) \log (P(Z \setminus U|d)p(z|d,Z \setminus U))dz$$

$$= -P(U|d) \log P(U|d) - P(Z \setminus U|d) \log P(Z \setminus U|d)$$

$$- P(U|d) \int_U p(z|d,U) \log p(z|d,U)dz$$

$$- P(Z \setminus U|d) \int_{Z \setminus U} p(z|d,Z \setminus U) \log p(z|d,Z \setminus U)dz$$  \hspace{1cm} (23)

$$= H[p_{\text{C}}(x|d)] + P(U|d)H[p(z|d,U)] + P(Z \setminus U|d)H[p(z|d,Z \setminus U)].$$  \hspace{1cm} (24)

\(^\text{1}\)Such improper distributions can be regarded as limits of proper distributions carrying only minimal information. Generally, the use of an improper distribution in Bayesian theory is without difficulties provided that the resulting posterior distribution constitutes again a proper distribution.
Lemma 1 will be used for the proof of the following theorem that states the explicit formula for \( p_{\text{ME}}(z|d) \).

**Theorem 1.** The optimization problem in Equation (18) has for \( z \in U \) the solution

\[
p_{\text{ME}}(z|d) = \begin{cases} 
\pi p(z|d,U), & \text{if } \beta \leq \pi , \\
\beta p(z|d,U), & \text{if } \pi < \beta ,
\end{cases}
\]

with \( \pi := \frac{2H[p(z|d,U)]}{2H[p(z|d,U)] + 2H[p_{\text{ME}}(z|d,Z\setminus U)]} \). \( (25) \)

Thereby, the logarithms in the entropy functions are taken to the base 2 and \( H[p_{\text{ME}}(z|d,Z \setminus U)] \) denotes the entropy of the non-informative Maximum Entropy distribution on \( Z \setminus U \). Furthermore, \( p_{\text{ME}}(z|d) \) must be also non-informative with regard to \( Z \setminus U \).

It should be noted that, in the formula for \( p_{\text{ME}}(z|d) \) in Equation (25), the factors \( \pi \) and \( \beta \), respectively, correspond to the posterior probability mass \( P(U|d) \) which the Maximum Entropy Principle assigns to the local context \( U \) (compare also Equation (7)).

**Proof of Theorem 1.** Obviously, maximizing the entropy \( H[p(z|d)] \) is equivalent to minimizing its negative counterpart \( -H[p(z|d)] \).

Using the abbreviation \( u := P(U|d) \), Lemma 1 states that

\[
-H[p(z|d)] = u \log u + (1 - u) \log(1 - u) - u H[p(z|d,U)] - (1 - u) H[p(z|d, Z \setminus U)]. \tag{26}
\]

Focussing on the local context \( U \) means ignoring its complement \( Z \setminus U \) completely. This means that absolutely no knowledge concerning \( Z \setminus U \) is available within the resulting focussed Bayesian model. Inspecting Equation (26) directly delivers that \( -H[p(z|d)] \) is minimal if \( H[p(z|d, Z \setminus U)] \) is chosen to be maximal. Hence, we must set

\[
p(z|d, Z \setminus U) = p_{\text{ME}}(z|d, Z \setminus U).
\]

The derivations made so far deliver the result that we must minimize the function

\[
f(u) := u \log u + (1 - u) \log(1 - u) - u H[p(z|d,U)] - (1 - u) H[p_{\text{ME}}(z|d, Z \setminus U)]
\]

under the constraint \( \beta \leq u \) that corresponds to Equation (13).

Calculating the second derivative of \( f(u) \) proves that \( f(u) \) is convex for \( u \in (0, 1) \).

The solution of such a convex optimization problem is based on considering the respective Lagrangian function\(^{26}\) which takes the form

\[
L(u, \lambda) = u \log u + (1 - u) \log(1 - u) - u H[p(z|d,U)] - (1 - u) H[p_{\text{ME}}(z|d, Z \setminus U)] + \lambda(\beta - u). \tag{28}
\]

The corresponding Karush-Kuhn-Tucker (KKT) conditions\(^{26}\) to be satisfied are:

\[
(a) \ \beta \leq u , \quad (b) \ \lambda \geq 0 , \quad (c) \ \lambda = 0 \lor \beta = u , \\
(d) \ \log \left( \frac{u}{1 - u} \right) - H[p(z|d,U)] + H[p_{\text{ME}}(z|d, Z \setminus U)] - \lambda = 0.
\]

A case distinction according to condition (c) is necessary at this point:

Firstly, if it holds \( \lambda = 0 \), condition (d) delivers

\[
\frac{u}{1 - u} = \frac{2H[p(z|d,U)] - H[p_{\text{ME}}(z|d,Z\setminus U)]}{2H[p(z|d,U)] + 2H[p_{\text{ME}}(z|d,Z\setminus U)]} \iff u = \frac{2H[p(z|d,U)]}{2H[p(z|d,U)] + 2H[p_{\text{ME}}(z|d,Z\setminus U)]} \tag{29}
\]

which holds according to condition (a) if \( \beta \leq u \).

Secondly, in the case \( \beta = u \), condition (d) delivers

\[
\lambda = \log \left( \frac{\beta}{1 - \beta} \right) - H[p(z|d,U)] + H[p_{\text{ME}}(z|d, Z \setminus U)]. \tag{30}
\]
Hence, according to condition (b), it must hold
\[ \frac{\beta}{1 - \beta} \geq 2^{\text{\footnotesize{\text{H}[p(z|d,U)] - H[p_{\text{ME}}(z|d,Z\setminus U)]}}} \quad \Leftrightarrow \quad \beta \geq \frac{2^{\text{\footnotesize{\text{H}[p(z|d,U)]}}} - 2^{\text{\footnotesize{\text{H}[p(z|d,U)]}}} + 2^{\text{\footnotesize{\text{H}[p_{\text{ME}}(z|d,Z\setminus U)]}}}}{2^{\text{\footnotesize{\text{H}[p(z|d,U)]}}} + 2^{\text{\footnotesize{\text{H}[p_{\text{ME}}(z|d,Z\setminus U)]}}}}. \] (31)

As described in Section 4.1, the Maximum Entropy Principle delivers that distribution possessing maximal uncertainty among all distributions being consistent with the given information. Having a closer look on the explicit formula for \( p_{\text{ME}}(z|d) \) in Equation (25) makes this fact apparent for the specific application case of the Maximum Entropy Principle which is considered in the current section.

Provided that \( \beta \leq \pi \), the Maximum Entropy Principle assigns the more global posterior probability mass \( P(U|d) \) on the local context \( U \), the more uncertain the local posterior probability \( p(z|d,U) \) is. Conversely, the more informative \( p(z|d,U) \) is, the less probable the Maximum Entropy Principle assumes the local context \( U \) to be globally seen.

In other words, the Maximum Entropy Principle assumes the worst case: obviously, the fact that the local posterior distribution is highly concentrated may be generally misleading. The transition from \( p(z|d) \) to \( p(z|d,U) \) which is done at the focussing leads to a distortion (in form of an overestimation) of probabilities within \( U \). The degree of this distortion is the higher, the lesser \( P(U|d) \) is (compare Equation (7)). The Maximum Entropy Principle rather assumes that a situation involving a high degree of distortion is on hand and tries to correct this high degree of distortion by flattening the global posterior distribution \( p(z|d) \) on \( U \) correspondingly much.

To our mind, the application of the Maximum Entropy Principle for the reduction of the probability interval scheme in Equation (14) to a unique posterior distribution is reasonable in a lot of cases as it sounds a note of caution when interpreting the results obtained by focussed Bayesian fusion. It is emphasized that this conclusion is made bearing in mind the fact that a meaningful bound \( \beta \) for \( P(U|d) \) and the corresponding condition \( \beta \leq \pi \) in Equation (25) may prevent the shifting of absurdly much probability mass on \( Z \setminus U \).

5. MAKING SUBSEQUENT DECISIONS

5.1 Decision theoretic concepts

The ultimate objective of most fusion tasks is to provide an adequate basis for further decision-making. Decisions to be reached subsequently to a fusion task may of various kinds. For example, a pedestrian participating in road traffic may (usually unconsciously) combine different kinds of information with the aim of reaching a decision whether or not to cross a busy road. To give another example, a technical system may support a human operator at fusing different kinds of information regarding the situation in a certain geographical area with the aim of reaching a decision whether or not to take certain measures for the protection of a critical infrastructure located within this area.

Bayesian theory is closely linked to statistical decision theory. By this, the results obtained by Bayesian fusion can be directly inserted into a founded and well-proven decision theoretic framework. If the posterior distribution \( p(z|d) \) has been obtained completely, the respective subsequent decision making is called decision making under risk. Key element of decision making under risk is the Principle of Expected Utility. Broadly speaking, according to the Principle of Expected Utility, a rational decision maker being confronted with different alternatives shall chose an alternative that is, in light of the present uncertainty, expected to be of most advantage for him.

More formally, denote by \( A \) the set of alternatives being available to the decision maker. At decision making under risk, it is assumed that the utility \( u(a,z) \) of each alternative \( a \in A \) depends on the “true” value \( z \) of the Properties of Interest which is not surely known. For example, the utility of taking or not taking certain measures for the protection of a critical infrastructure may depend on type, position and driving direction of objects being present in a certain geographical area. Bayesian fusion may be performed prior to the actual decision-making with the aim to reduce the existing uncertainty with regard to these three object characteristics. According to
the Bayesian paradigm, the posterior distribution \( p(z|d) \) then comprehensively represents the information being available about them after the fusion has been performed.

From a mathematical point of view, the Principle of Expected Utility states that a rational decision maker shall chose among the available alternatives contained in \( A \) an alternative \( a_{\text{opt}} \) such that it holds\(^ {11} \)

\[
a_{\text{opt}} = \arg \max_{a \in A} E_{p(z|d)} [u(a, z)] \quad \text{with} \quad E_{p(z|d)} [u(a, z)] = \int_{Z} u(a, z) p(z|d) dz .
\]

(32)

The motivation for the local consideration of fusion tasks in Section 3.1 directly makes clear that this standard framework of statistical decision theory that has been presented so far also corresponds to an idealization in a lot of real world tasks. In practice, the needed posterior probability distribution \( p(z|d) \) will often not be known exactly and completely as it is demanded within this framework. There is the general need to adopt the underlying decision theoretic concepts to cases in that only partial information regarding the involved (posterior) probability distribution is available.

Literature provides different criteria and approaches for such decision making under partial information. Especially notable in light of this publication is the Lazy Decision Making approach.\(^ {29} \) Being based on the theory of Linear Partial Information,\(^ {30} \) the approach provides good opportunities for a theoretical derivation of the quantities to be considered. Also, the approach provides founded concepts for the purposeful precision of partial information in cases in that it is not possible to come to a decision being sufficiently well grounded.

5.2 Further analysis of partial decision making in the context of focussed Bayesian fusion

The partial information with regard to the global posterior distribution \( p(z|d) \) being obtained via focussed Bayesian fusion corresponds to the knowledge that \( p(z|d) \) must be consistent with Equation (12) and Equation (14) (or, alternatively, with Equation (7) and Equation (13)). Choosing a unique posterior distribution among the set \( \Pi \) of all probability distributions satisfying these constraints as done in Section 4 is a special kind of a decision.\(^ {11} \) Obviously, it is not sensible to make such a decision if the fusion result (i.e., the posterior distribution) shall be subsequently inserted into a decision theoretic framework to chose between different alternatives. By this, two hard decisions would succeed.

In addition, the probability interval scheme given in Equation (14) represents (together with Equation (12)) the uncertainty resulting due to the focussing on the local context \( U \) in a separate manner (i.e., separated from the uncertainty that would still remain with regard to \( z \) if the Bayesian fusion was performed with regard to the whole range of values of the Properties of Interest). Generally, enabling the decision maker to analyze how exactly the focussing on the local context \( U \) influences his basis for decision-making will be promising. Especially, on basis of this analysis, the decision maker may decide to improve his current basis for decision-making by complementing the partial information being available with regard to \( p(z|d) \) via an enlargement of the local context \( U \). For providing these options to the decision maker, the described separation of uncertainties is indispensable.

The concepts of decision making under partial information provide precisely the means for respecting this separation in the decision making task. Instead of one unique posterior distribution, the set of all possible posterior distributions \( \Pi \) is inserted in the decision theoretic framework, here. As consequence, instead of one unique value for the expected utility of each alternative \( a \in A \), a set \( \Gamma_a \) of possible values for the expected utility of this alternative results:

\[
\Gamma_a = \{ E_{p(z|d)} [u(a, z)] \mid p(z|d) \in \Pi \} .
\]

(33)

Obtaining knowledge with regard to certain characteristics of the sets \( \Gamma_a, a \in A \), as well as with regard to the interrelation of their elements enables the decision maker to reach a founded conclusion whether decision making on basis of the available partial information is acceptable for him and (if yes) what alternative \( a \in A \) he shall chose.

\(^ {11} \) In this publication, it is assumed that all used utility functions are bounded and that an alternative maximizing the expected utility exists.
The concept of dominance of alternatives plays a central role in this regard. A certain alternative \(a^* \in A\) is defined to be dominating another alternative \(a^{**} \in A\) if it holds that

\[
E_{p(z|d)}[u(a^{**}, z)] \leq E_{p(z|d)}[u(a^*, z)] \quad \text{for all } p(z|d) \in \Pi.
\]

Equation (34) means that the given (partial) information is sufficient to reach the conclusion that \(a^*\) is surely to be preferred over \(a^{**}\) (in terms of expected utility). A sufficient (but not necessary) criterion for this is that the inequality \(\max \Gamma_{a^{**}} \leq \min \Gamma_{a^*}\) holds. Using the concept of dominance, the decision maker will often be able to discard certain alternatives being surely not optimal in terms of Equation (32). Especially, in the case that a particular alternative dominates all the alternatives being contained in the set \(A\) of possible alternatives, it will be even clear to him that this certain alternative is surely an optimal one.

Another central concept in context of decision making under partial information is the concept of regret. The regret \(R(a)\) of each alternative \(a \in A\) is defined as the maximal deficit (in terms of expected utility) that can arise from the choice of \(a\), i.e.,

\[
R(a) := \max_{a^* \in A} \max_{p(z|d) \in \Pi} \{E_{p(z|d)}[u(a^*, z)] - E_{p(z|d)}[u(a, z)]\}.
\]

By considering the regret values of the available alternatives, the decision maker may be able to choose an alternative which may be really good enough with respect to the task at hand – although being not granted to be an optimal one in terms of Equation (32). Similarly, in the case in which no optimal alternative in terms of Equation (32) is identifiable and the available resources of storage and computation capacities do not permit an actually needed subsequent improvement of the basis for decision making, the consideration of the regret values of the available alternatives will enable the decision maker to make at least the best decision being possible under these circumstances. Note that, to be able to interpret regret values correctly, the decision maker should also have the possibility to access the sets \(\Gamma_a\) themselves appropriately.

Of course, the practical application of the described concepts for decision making under partial information requires that the involved quantities and sets, respectively, are efficiently computable or at least estimable. Until today, the most feasible techniques for decision making under partial information deal only with quantities on finite sets. The application of such techniques to arbitrary decision problems is still an actual research theme.\(^3\)\(^1\)\(^2\)

In the context of focussed Bayesian fusion, the respective decision problem under partial information is rather unproblematically traceable back to the case of finite sets by imposing a certain restriction. This restriction is a minor one as it applies only to the set \(Z \setminus U\), not to the local context \(U\). More concretely, it is assumed that \(Z \setminus U\) corresponds (possibly after discretization) to a finite set. By this assumption, the given decision problem can be even traced back to a decision problem under Linear Partial Information which is addressable in a very uncomplicated analytic manner.\(^2\)

Partial information is called Linear Partial Information if the set of probability distributions being consistent with this information can be described via linear inequalities.\(^10\) It is common knowledge that by taking geometrical considerations into account, linear optimization on such a set is realizable very efficiently. More concretely, such a set constitutes a convex polyhedron and the search for a solution of a linear optimization problem can be restricted on the set of edges of the convex polyhedron.\(^33\)\(^34\)

In a previous publication of one of the authors,\(^2\) explicit formulas for the quantities and sets, respectively, being involved in the presented central concepts for decision making under partial information have been derived in a formal mathematical manner. Therefore, all needed derivations have been traced back to linear optimization.

According to the obtained results, the sets \(\Gamma_a\) of possible expected utility values for alternatives \(a \in A\) correspond to intervals of the form \(\Gamma_a = [\min \Gamma_a, \max \Gamma_a]\) with

\[
\min \Gamma_a = \min \left\{ (1 - \beta) \min_{z \in Z \setminus U} u_\Delta(a, z) + \beta E_{p(z|d,U)}[u(a, z)], E_{p(z|d,U)}[u(a, z)] \right\}, \tag{36}
\]

\[
\max \Gamma_a = \max \left\{ (1 - \beta) \max_{z \in Z \setminus U} u_\Delta(a, z) + \beta E_{p(z|d,U)}[u(a, z)], E_{p(z|d,U)}[u(a, z)] \right\}. \tag{37}
\]
Thereby, \( u_\Delta(a, z) \) denotes a suitably discretized version of the utility function \( u(a, z) \) within \( \mathbb{Z} \setminus \mathbb{U} \).

Furthermore, also an additional dominance criterion (usable to detect dominance relations in cases in which the expected utility intervals of the considered alternatives overlap) and an upper bound for the regret values of the available alternatives have been derived.\(^2\) We consciously refrain from citing the respective formulas in detail, here, as the details are not essential for the fundamental understanding of the resulting decision making techniques. However, it is important to emphasize, that the explicit calculation of the derived quantities is in essence possible on basis of the knowledge being available at focussed Bayesian fusion (which gets directly clear with regard to the calculation of the expected utility intervals from an inspection of Equation (36) and Equation (37)).

6. CONCLUSION

In the present contribution, local Bayesian fusion and especially focussed Bayesian fusion have been motivated and analysed. Basic considerations, why a local concentration of fusion tasks makes sense and what the resulting effects of the local handling are, have been exposed. The idea of focussed Bayesian fusion as a possible way to realize local Bayesian fusion by restricting the calculation totally to a local context has been introduced.

In order to derive a unique posterior distribution for the whole domain of the Properties of Interest from focussed Bayesian fusion with as much significance as possible, the Maximum Entropy Principle has been shown to be reasonably applicable. Therefore, the Maximum Entropy Principle and the Principle of Minimum Information have been introduced. The results of the application of these principles to the task of focussed Bayesian fusion have then been derived and analyzed. In consequence, it can be concluded that applying the Maximum Entropy Principle leads to a worst case estimation, i. e., a very careful outcome of the fusion methodology. Even if a cautious interpretation of the result obtained by focussed Bayesian fusion is certainly appropriate, this conservatism may, however, in principle lead to posterior distributions which are too flat within the local context and which have, in consequence, too little significance. Nonetheless, we are convinced that the proposed cautious consideration is advantageous, since this procedure may guarantee that the significance of the results obtained by focussed Bayesian fusion is not overestimated.

As a further key aspect of this contribution, we investigated how focussed Bayesian fusion can be integrated directly in statistical decision making. In this context, it can be stated that the necessity to model local views exceeds the topic of fusion and concerns also decision theoretic concepts. Thus, approaches from decision theory establish a sensible foundation. Here, the concepts of decision making under partial information can be applied to deduce the optimal alternative. That way, it is possible to use the partial information generated by focussed Bayesian fusion directly in the decision making process without the preceding step of determining a unique posterior distribution. In more detail, tracing back the given problem to a problem under Linear Partial Information allows to use the respective concepts efficiently.

As an outlook, the presented approaches of local and focussed Bayesian fusion may be combined with the concepts of conjugate distributions or Markov Chain Monte Carlo (MCMC) methods. Since these methods are well-proven ways to reduce high computational costs of Bayesian fusion, it can be expected that incorporating local or focussed Bayesian fusion approaches may offer additional promising options.

REFERENCES


