Belief Revision and Information Fusion in a Probabilistic Environment

Gabriele Kern-Isberner
FernUniversität Hagen
Dept. of Computer Science
D-58084 Hagen, Germany
gabriele.kern-isberner@fernuni-hagen.de

Wilhelm Rödder
FernUniversität Hagen
Dept. of Economics
D-58084 Hagen, Germany
wilhelm.roedder@fernuni-hagen.de

Abstract

This paper presents new methods for probabilistic belief revision and information fusion. By making use of the principles of optimum entropy (ME-principles), we define a *generalized revision operator* which aims at simulating the human learning of lessons, and we introduce a fusion operator which handles probabilistic information faithfully. In general, this fusion operator computes kind of mean probabilistic values from pieces of information provided by different sources.

Introduction

Most of the knowledge which is used e.g. in advanced knowledge base systems, or in cognitive modelling is uncertain, incomplete, and subject to changes. A formal concept for knowledge management should not only focus on reasoning techniques, but also make methods for revising and updating on new information available, and allow for the fusing of knowledge from different sources. Only such a combination of methods makes sure that the resulting knowledge base or belief state, respectively, yields optimal results. To realize such an ambitious concept, an expressive and powerful framework with solid theoretical foundations and sophisticated computation techniques is needed.

Probability theory, the oldest approach to handle uncertain knowledge, meets all these requirements. Instead of only two extreme truth values, as in classical frameworks, a continuum of numerical values between 0 (*false*) and 1 (*true*) is available to specify degrees of belief. Probability theory was also the first framework to address explicitly the problem of belief change, offering with Bayesian conditioning and Jeffrey's conditioning elegant methods to adjust one's stock of belief to certain or uncertain evidence, respectively. The abundance of probabilistic models (i.e. distributions), however, makes inference from incomplete knowledge bases quite weak. Here, the use of information theoretical techniques is apt to improve probabilistic reasoning substantially. The principle of maximum entropy (Jaynes 1983; Paris 1994) represents given probabilistic knowledge inductively by a "best" model, i.e. without adding information unnecessarily, which can be used for inferences (*ME-representation*). Its generalization, the principle of minimum cross-entropy (Shore & Johnson 1980; Kern-Isberner 1998), provides an optimal method to revise a given prior distribution by new information, generalizing both Bayesian and Jeffrey's conditioning (*ME-revision*). In (Rödder & Xu 1999), first steps towards making use of this paradigm for fusing information have been taken.

Copyright © 2003, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

In this paper, we elaborate further on these topics. First, we will outline how to use the ME-revision method to realize different belief change operations, generally known as *(genuine) revision*, *updating*, and *focusing*. Then, we will present a new probabilistic revision operator which is intermediate between revision and updating, and aims at simulating the human learning of lessons. The basic idea here is the following: Even if previously learned lessons are erroneous in some respect, further experiences may correct the mistakes, without giving up completely the prior knowledge. Indeed, the prior knowledge can be recovered from the resulting posterior belief state. In particular, the lessons to be learned need not be compatible with one another. We will illustrate by an example that the new operator is able to revise even certain prior beliefs which is not possible by the standard ME-approach.

The second main point of this paper is how to use the ME-methodology for fusing probabilistic beliefs. We take up the idea from (Rödder & Xu 1999), but show that a naïve approach yields undesired, weakening effects. We fix this problem and define an *ME-fusion-operator* that computes kind of mean value from probabilities describing degrees of belief provided by different experts. In particular, if all experts agree on that probability, ME-fusion returns this same probability. So, the fusion method to be presented here handles information faithfully, without allowing weakening or reinforcing effects.

This paper is organized as follows: We start with describing the formal logical background of this paper, and recalling some basic facts on the principles of optimum entropy. The following section deals with probabilistic belief revision and presents a *generalized revision operator*. Afterwards, we describe our ME-fusion technique. We conclude with highlighting the main contributions of this paper, and give an outlook on further work.

Formal background and notations

Let $\mathcal{L}$ be a propositional logical language, finitely generated by propositional variables from $\mathcal{V} = \{a, b, c, \ldots\}$. Formulas from $\mathcal{L}$ are well-formed in the usual way by use of the junctors $\wedge$ (*and*), $\vee$ (*or*), and $\neg$ (*not*), and will be denoted by uppercase letters, $A, B, C, \ldots$. To simplify notation, we will omit the $\wedge$-junctor in conjunctions and simply juxtapose conjuncts, i.e. $AB = A \wedge B$. Moreover, instead of $\neg A$, we will write $\overline{A}$. Ω is the set of all classical-logical interpretations of $\mathcal{L}$. This Boolean frame is extended by a binary, non-Boolean conditional operator, $|$. Formulas of the form $(B|A)$ are called *conditionals*, or simply *rules*. In a second extension step, propositional and conditional formulas are assigned a real number $x \in [0, 1]$, representing a probability. Propositional formulas $A \in \mathcal{L}$ are identified with the conditional $(A|\top)$, where $\top$ is any tautology. So, the syntactical objects we will consider in this paper have the form $(B|A)[x]$, $A, B \in \mathcal{L}, x \in [0, 1]$.

As to the semantics, the models are probability distributions, P, over the propositional variables $\mathcal{V}$. P *satisfies a conditional* $(B|A)[x]$, $P \models (B|A)[x]$, iff $P(A) > 0$ and $P(B|A) = \frac{P(AB)}{P(A)} = x$. This satisfaction relation generalizes to sets of conditionals in a straightforward way. Our probabilities will be subjective probabilities, so the elements $\omega \in \Omega$ should be taken as possible worlds, rather than statistical elementary events. A set $\mathcal{R} = \{(B_1|A_1)[x_1], \ldots, (B_m|A_m)[x_m]\}$ of probabilistic conditionals is *consistent* iff it has a model P, i.e. iff there is a distribution P with $P \models \mathcal{R}$. Otherwise, it is called *inconsistent*. Given a distribution P, $\mathcal{R}$ is called *P-consistent* iff there is a model $Q \models \mathcal{R}$ such that Q is totally continous with respect to P (i.e. $P(\omega) = 0$ implies $Q(\omega) = 0$). For a set $\mathcal{R} = \{(B_1|A_1)[x_1], \ldots, (B_m|A_m)[x_m]\}$ of probabilistic conditionals and a propositional formula W, we define the *conditioning of* $\mathcal{R}$ *by* W via

$$\mathcal{R}|W := \{(B_1|A_1W)[x_1], \ldots, (B_m|A_mW)[x_m]\}$$

Probabilistic reasoning on optimum entropy

The principles of maximum entropy and of minimum cross-entropy are well-known techniques to represent (mostly incomplete) probabilistic knowledge inductively, and to combine new probabilistic information with prior knowledge, respectively (cf. e.g. (Jaynes 1983)). Entropy and cross-entropy are notions stemming from information theory. Entropy quantifies the indeterminateness inherent to a distribution P by $H(P) = -\sum_\omega P(\omega) \log P(\omega)$, and cross-entropy measures the information-theoretical distance of a distribution P with respect to a distribution Q by $R(Q, P) = \sum_{\omega \in \Omega} Q(\omega) \log \frac{Q(\omega)}{P(\omega)}$.

Given a set $\mathcal{R}$ of probabilistic conditionals, the *principle of maximum entropy*

$$\max H(Q) = -\sum_{\omega} Q(\omega) \log Q(\omega) \qquad (1)$$

s.t. Q is a probability distribution with $Q \models \mathcal{R}$.

solves the problem of representing $\mathcal{R}$ by a probability distribution without adding information unnecessarily. The resulting distribution is denoted by *ME*($\mathcal{R}$). The *principle of minimum cross-entropy* allows us to take prior knowledge, represented by a distribution P, also into account. It yields a solution to the following problem: How should P be modified to obtain a (posterior) distribution P^* with $P^* \models \mathcal{R}$ by solving the minimization problem

$$\min R(Q, P) = \sum_{\omega \in \Omega} Q(\omega) \log \frac{Q(\omega)}{P(\omega)} \qquad (2)$$

s.t. Q is a probability distribution with $Q \models \mathcal{R}$

Both problems are solvable, provided that $\mathcal{R}$ is consistent, or P-consistent, respectively (Csiszár 1975). The principle of minimum cross-entropy can be regarded as more general than the *principle of maximum entropy*. Indeed, maximizing (absolute) entropy under some given constraints $\mathcal{R}$ is equivalent to minimizing cross-entropy with respect to the uniform distribution, given $\mathcal{R}$. Therefore, we refer to both principles as the *ME-principle*, where the abbreviation *ME* stands both for *M*inimum cross-*E*ntropy and for *M*aximum *E*ntropy. Since both approaches follow the paradigm of *informational economy*, they are particularly well-behaved, as has been proved by different authors (Shore & Johnson 1980; Paris 1994; Kern-Isberner 1998; 2001); for more detailed information on these techniques, we refer the reader to this literature.

All ME-calculations in this paper have been carried out with the system shell SPIRIT, which is available via http://www.fernuni-hagen.de/BWLOR/forsch.html. For more detailed information on computational issues, cf. (Rödder & Meyer 1996).

Revising probabilistic beliefs

The ME-principle yields a probabilistic belief revision operator, $*_{ME}$, associating to each probability distribution P (representing a probabilistic belief state) and each P-consistent set $\mathcal{R}$ of probabilistic conditionals a revised distribution $P_{ME} = P *_{ME} \mathcal{R}$ in which $\mathcal{R}$ holds. Classical belief revision theory (Alchourrón, Gärdenfors, & Makinson 1985; Gärdenfors 1988) distinguishes between different kinds of belief change: *(genuine) revision* takes place when new information about a static world arrives, whereas *updating* tries to incorporate new information about a (possibly) evolving, changing world (Katsuno & Mendelzon 1991). *Focusing* (Dubois & Prade 1997) means applying generic knowledge to the evidence present by choosing an appropriate context or reference class. All these different facets of belief change can be realized via the ME-approach, as follows:

- In its original form, the ME-operator $*_{ME}$ corresponds to *updating* – a complete set of beliefs about a prior world which is not assumed to be static is adapted to new probabilistic beliefs. None of the prior beliefs can be expected still to hold in the new belief state. So, updating is kind of *successive* belief change.
- A typical situation where genuine revision can be carried out is the following: Our explicit knowledge about a specific world is given by a set $\mathcal{R}$ of probabilistic conditionals, and some new information, $\mathcal{S}$, on that same world arrives. Since $\mathcal{R}$ and $\mathcal{S}$ refer to the same world, $\mathcal{R} \cup \mathcal{S}$ must be consistent (erroneous information is excluded here).

Mostly, also background knowledge is available which is assumed to be represented by a probability distribution, P. The prior belief state before coming to know $\mathcal{S}$ then is $P *_{ME} \mathcal{R}$, combining background and situational knowledge. Note that although $\mathcal{S}$ is assumed to be consistent with $\mathcal{R}$, it will be incompatible with P in general, in that $P \not\models \mathcal{S}$. The *(genuine) revision* of $P *_{ME} \mathcal{R}$ then is defined by $P *_{ME} (\mathcal{R} \cup \mathcal{S})$. Prior knowledge $\mathcal{R}$ and new information $\mathcal{S}$ are taken to be on the same level, and revision is done by kind of *simultaneous* adoption of both to background knowledge.

- In a probabilistic setting, *focusing* on a certain evidence, A, is usually done by conditioning. Since $P(\cdot|A) = P *_{ME} \{A[1]\}$, the ME-revision operator offers a straightforward way to realize focusing on uncertain evidence $A[x]$, $x \in (0,1)$, by forming $P *_{ME} \{A[x]\}$. Hence, ME-focusing amounts to updating with certain or uncertain facts.

The formal properties of these revision operators have been investigated in (Kern-Isberner 2001) and related to those of classical belief revision operators. Note that indeed, ME-revision and ME-updating are different, since $(P *_{ME} \mathcal{R}) *_{ME} \mathcal{S} \neq P *_{ME} (\mathcal{R} \cup \mathcal{S})$ in general.

So different forms of belief revision can be realized by making use of one methodology, but doing so in different ways. This sheds a new, unifying light on the belief change framework – usually, the different change operators are considered as substantially different, a view that is hardly compatible with the ease with which humans adapt their stock of belief to new situations.

With the help of the ME-approach, however, also other, intermediate forms of belief change can be realized. In the following, we will present a method that is apt to perform a revision operation in case that prior and new knowledge are *not* compatible, since the prior knowledge is erroneous. This new method follows the line of considering human learning as adapting knowledge in packages or lessons. New knowledge may override or refine old experiences, but the prior knowledge is still present in the new belief state to a certain degree. The corresponding revision operator will be denoted by $\diamond$.

Suppose our prior knowledge is described by a set $\mathcal{R}$ of probabilistic rules, and a package of new information is represented by another set $\mathcal{S}$ of probabilistic conditionals. $\mathcal{R}$ and $\mathcal{S}$ need not be compatible, so $\mathcal{R} \cup \mathcal{S}$ may be inconsistent. We set priority to $\mathcal{S}$, taking the new information for granted, so the prior knowledge $\mathcal{R}$ has to be weakened to make a combination of $\mathcal{R}$ and $\mathcal{S}$ on the same level possible. The prior belief state to be revised is $P_1 := ME(\mathcal{R})$.

Let E_1 stand for the first experience where we learnt $\mathcal{R}$. The degree to which $\mathcal{R}$ should be present in the new belief state is specified by $x \in [0,1]$. The *generalized revision of* $P_1 = ME(\mathcal{R})$ *by* $\mathcal{S}$ is defined by

$$P_1 \diamond \mathcal{S} := ME(\mathcal{R}|E_1 \cup \mathcal{S} \cup \{E_1[x]\}) \qquad (3)$$

$P_1 \diamond \mathcal{S}$ represents the new information, $\mathcal{S}$, while the prior knowledge, $\mathcal{R}$, can be recovered from $P_1 \diamond \mathcal{S}$ by conditioning on E_1. This generalized type of belief revision is illustrated by the following example.

Example 1 Grete, as a young child, learned about birds' abilities to fly, sing, run, dive, and swim. This knowledge, $\mathcal{R}_1$, was very general, and there was no specification concerning different kinds of birds, like penguins, ostriches, craws, song-birds, or aquatic birds:

Grete's young child lesson, $\mathcal{R}_1$:

conditional	prob.	conditional	prob.
(*sing* | *bird*)	0.80	(*fly* | *bird*)	1.00
(*run* | *bird*)	0.10	(*dive* | *bird*)	0.01
(*swim* | *bird*)	0.00		

The belief state corresponding to Grete's young age is given by $ME(\mathcal{R}_1)$ in which all rules of $\mathcal{R}_1$ are present. Note that the conditional relationships have been learned in a directed way, so we can not expect, for instance, the conditional (*bird* | *fly*) to have a high probability in $ME(\mathcal{R}_1)$, too. Indeed, we find $ME(\mathcal{R}_1)(bird \mid fly) = 0.1311$, so, flying objects are not assumed by default to be birds. However, a positive correlation between *birds* and *fly* has been established, since the probability that an object is a bird ($ME(\mathcal{R}_1)(bird) = 0.0702$) can be increased considerably by observing its flying. The low probability of being a *bird* corresponds to the high informativeness that has been attached to birds by the ME-approach (for the connection between conditionals, probability, and information, cf. (Rödder & Kern-Isberner 2003)).

Later, when Grete grew up, her knowledge increased, and was refined, as is shown in rule set $\mathcal{R}_2$; in particular, she came to know birds that were not able to fly but excellent swimmers and divers, which is in conflict with her prior knowledge:

Grete's child lesson, $\mathcal{R}_2$:

conditional	prob.	conditional	prob.
(*sing* | *songbird*)	0.99	(*dive* | *songbird*)	0.05
(*sing* | *aquab*)	0.01	(*dive* | *aquab*)	0.50
(*sing* | *penguin*)	0.00	(*dive* | *penguin*)	0.99
(*sing* | *ostrich*)	0.00	(*dive* | *ostrich*)	0.01
(*sing* | *craw*)	0.01	(*dive* | *craw*)	0.05
(*fly* | *songbird*)	0.99	(*swim* | *songbird*)	0.05
(*fly* | *aquab*)	0.96	(*swim* | *aquab*)	0.99
(*fly* | *penguin*)	0.01	(*swim* | *penguin*)	0.99
(*fly* | *ostrich*)	0.01	(*swim* | *ostrich*)	0.05
(*fly* | *craw*)	1.00	(*swim* | *craw*)	0.05
(*run* | *songbird*)	0.05	(*bird* | *songbird*)	1.00
(*run* | *aquab*)	0.10	(*bird* | *aquab*)	1.00
(*run* | *penguin*)	0.05	(*bird* | *penguin*)	1.00
(*run* | *ostrich*)	0.99	(*bird* | *ostrich*)	1.00
(*run* | *craw*)	0.99	(*bird* | *craw*)	1.00

Let $P_2 = ME(\mathcal{R}_1) \diamond \mathcal{R}_2$ be the result of revising $ME(\mathcal{R}_1)$ by $\mathcal{R}_2$ in a generalized way, as defined in (3), with $x = 0.9$. In particular, Grete revised her knowledge from $\mathcal{R}_1$, as follows:

$P_2(sing \mid bird)$	= 0.68		$P_2(fly \mid bird)$	=	0.85
$P_2(run \mid bird)$	= 0.22		$P_2(dive \mid bird)$	=	0.17
$P_2(swim \mid bird) = 0.18$					

Note that no new general relationships between the superclass *bird* and the above properties are explicitly stated in

$\mathcal{R}_2$; rather, the probabilities of the conditionals in $\mathcal{R}_1$ have been revised implicitly by observing counterexamples. ■

The degree x to which earlier experience should be present in the current belief state is part of the modelling and should be set appropriately by the user. The case where x is chosen as close to 1 as possible is discussed in (Reucher & Rödder 2002).

Information fusion via the ME-approach

When building up knowledge bases, one often is presented with the problem to fuse knowledge coming from different experts, or to combine expert knowledge with knowledge from statistical data. In general, *information fusion* is concerned with merging information that stems from different sources in order to make this merged information usable for purposes such as query answering, or decision making (for a profound discussion of this notion and a broad overview, cf. (Bloch, Hunter, & others 2001)). In the framework dealt with in this paper, information is provided by sets $\mathcal{R}_1, \ldots, \mathcal{R}_n$ of probabilistic conditionals. We assume each of these sets $\mathcal{R}_i$ to be consistent; inconsistent rule bases have to be split up into consistent subbases. The union $\cup_{i=1}^{n} \mathcal{R}_i$, however, may be inconsistent, or even plainly contradictory in that identical facts or conditionals may be assigned different probabilities. In the following, we will explain how to use ME-methodology for information fusion in this case; in detail, we will

- merge the rule bases $\mathcal{R}_1, \ldots, \mathcal{R}_n$ into one (consistent) rule base $\mathcal{R} := \mathcal{R}_1 \odot \ldots \odot \mathcal{R}_n$, and then
- use the principle of maximum entropy to build up a complete probability distribution $F_{ME}(\mathcal{R}_1, \ldots, \mathcal{R}_n) = \mathit{ME}(\mathcal{R}_1 \odot \ldots \odot \mathcal{R}_n)$ which can be used for further inferences.

Note that, although the terms *merging* and *fusion* are often used synonymously, in our two-step approach we will use *merging* for the combination of *knowledge bases* $\mathcal{R}_1, \ldots, \mathcal{R}_n$, and *fusion* for the overall aggregation of knowledge in one (ideal) belief state $F(\mathcal{R}_1, \ldots, \mathcal{R}_n)$. The degrees of belief reflected by $F(\mathcal{R}_1, \ldots, \mathcal{R}_n)$ should arise as a compromise of the degrees of belief provided by each agent.

The basic idea is to consider each set $\mathcal{R}_i$ of probabilistic rules as describing the world from a certain point of view, W_i, provided by an intelligent agent i, and to condition the rules accordingly. This eliminates inconsistencies between the rule bases. So, for each agent $i, 1 \leqslant i \leqslant n$, a new binary variable W_i is introduced representing their point of view. Then the union $\mathcal{R}_1|W_1 \cup \ldots \cup \mathcal{R}_n|W_n$ is consistent. The following simple example shows, however, that this first, naïve approach to merge probabilistic rule bases yields undesired effects when combined with ME-methodology.

Example 2 We consider the following sets $\mathcal{R}_1, \mathcal{R}_2$ of probabilistic rules specified by two different experts: $\mathcal{R}_1 = \{A[0.7]\}$, $\mathcal{R}_2 = \{A[0.8]\}$. The experts disagree on the probability of the fact A, so $\mathcal{R}_1 \cup \mathcal{R}_2$ is obviously inconsistent. Conditioning each fact on the corresponding expert, however, makes both pieces of knowledge compatible – $\mathcal{R}_1|W_1 \cup \mathcal{R}_2|W_2$ is consistent. Thus, $P^* := \mathit{ME}(\mathcal{R}_1|W_1 \cup \mathcal{R}_2|W_2)$ can be computed, and $P^*(A)$ is to reflect the fused information. But we find $P^*(A) = 0.6613$ – a disappointingly low value. We might have expected kind of average between 0.7 and 0.8, the information, however, got weakened. On the other hand, relating explicitly to both experts yields a reinforcing effect: $P^*(A|W_1W_2) = 0.8270$, a value which is likewise not within the interval $[0.7, 0.8]$.

These unexpected effects become even more evident when we assume that both experts agree and specify the same degree of belief for A: $\mathcal{R}_1{}' = \mathcal{R}_2{}' = \{A[0.7]\}$. Here, the most intuitive result of a fusion process would be to assign A the probability 0.7. Constructing a distribution $P^{*\prime} := \mathit{ME}(\mathcal{R}_1{}'|W_1 \cup \mathcal{R}_2{}'|W_2)$ in the same way as above, however, yields the probabilities $P^{*\prime}(A) = 0.6325$ and $P^{*\prime}(A|W_1W_2) = 0.7598$. ■

Although there may be good reasons for such weakening or reinforcing effects, in general, each knowledge base $\mathcal{R}_i$ should be taken as an independent piece of information. Our main focus is on a proper definition of the merging operator $\odot$ to combine rule bases $\mathcal{R}_1, \ldots \mathcal{R}_n$. As outlined above, the actual fusion work will be done by ME-technology in a straightforward way.

The problems in Example 2 arise from unwanted interactions between the (different) sources of knowledge. So, further probabilistic information has to be added to ensure that the W_i provide a comprehensive, non-interfering views of the world. We define knowledge base merging by

$$\mathcal{R}_1 \odot \ldots \odot \mathcal{R}_n := \mathcal{R}_1|W_1 \cup \ldots \cup \mathcal{R}_n|W_n \cup \{W_1 \vee \ldots \vee W_n[1], W_iW_j[0], 1 \leqslant i,j \leqslant n, i \neq j\} \quad (4)$$

The additional information has to be interpreted within the ME-framework: $W_1 \vee \ldots \vee W_n[1]$ ensures that no information from the outside is allowed to intrude into the resulting distribution, and $W_iW_j[0]$ for $i \neq j$ precludes interferences between information coming from different agents. The ME-fusion operation is now realized in a straightforward way:

$$F_{ME}(\mathcal{R}_1, \ldots, \mathcal{R}_n) := \mathit{ME}(\mathcal{R}_1 \odot \ldots \odot \mathcal{R}_n) \quad (5)$$

$F_{ME}(\mathcal{R}_1, \ldots, \mathcal{R}_n)$ yields the desired ideal probabilistic belief state representing the fused pieces of information.

First, we check whether the so-defined fusion operation yields more intuitive results in Example 2.

Example 3 Let $\mathcal{R}_1$ and $\mathcal{R}_2$ be as in Example 2, and let $P_1^* := F_{ME}(\mathcal{R}_1, \mathcal{R}_2)$. Now the fused information concerning A is computed as $P_1^*(A) = 0.7472$, indeed kind of mean value between 0.7 and 0.8. Let us consider again the case that both experts agree, i.e. $\mathcal{R}_1 = \mathcal{R}_2 = \{A[0.7]\}$. ME-fusion, as defined in (5), yields $F_{ME}(\mathcal{R}_1, \mathcal{R}_2)(A) = 0.7$, as desired. ■

The following example illustrates the method in a more complex case.

Example 4 Two physicians argue about the relevance of a symptom, A, for diseases B, C, D. They both agree that A is a good indicator of disease B, although they disagree on its estimated degree of relevance: One physician specifies his corresponding belief as $(B|A)[0.9]$, whereas the other physician considers $(B|A)[0.8]$ more appropriate. In case that B can definitely excluded, however, the first physician holds strong belief in C ($(C|A\overline{B})[0.8]$), whereas the second physician thinks D to be the most probable diagnosis ($(D|A\overline{B})[0.9]$). So, the two packages of information to be fused are $\mathcal{R}_1 = \{(B|A)[0.9], (C|A\overline{B})[0.8]\}$ and $\mathcal{R}_2 = \{(B|A)[0.8], (D|A\overline{B})[0.9]\}$, and the final result of the fusion process is represented by $P^* := \mathit{ME}(\mathcal{R}_1 \odot \mathcal{R}_2)$. From P^*, we obtain the following probabilistic answers for the queries listed below:

query	prob.	query	prob.
$(B\|A)$	0.85	$(D\|A\overline{B}W_1)$	0.50
$(C\|A)$	0.51	$(C\|A\overline{B}W_2)$	0.50
$(D\|A)$	0.54	$(C\|A\overline{B})$	0.59
		$(D\|A\overline{B})$	0.78

This table shows that although both experts favor one of C and D in case that $A\overline{B}$ is present, no unjustified bias as to the respective other diagnosis has been introduced by the ME-approach ($P^*(D|A\overline{B}W_1) = P^*(C|A\overline{B}W_2) = 0.50$). If only A is known, then P^* reflects the expected high probability for disease B ($P^*(B|A) = 0.85$), whereas the probabilistic belief in C or D, respectively, is quite low ($P^*(C|A) = 0.51$, $P^*(D|A) = 0.54$). The probabilities attached to $(C|A\overline{B})$ and $(D|A\overline{B})$ can be used to find a proper diagnose if symptom A is present, but diagnose B can be excluded. In this case, a clear vote for diagnose D can be derived from the fused knowledge of both experts ($P^*(D|A\overline{B}) = 0.78$ vs. $P^*(C|A\overline{B}) = 0.59$). This can be explained as follows: The first physician establishes quite a strong connection between A and B by stating $(B|A)[0.9]$. This connection gets lost when it becomes obvious that A, but not B, is present. In that case, diagnosis D is assigned the higher probability, so $P^*(D|A\overline{B}) > P^*(C|A\overline{B})$ should be expected. The significant difference in both probabilities can be attributed to a bias towards the second physician ($P^*(W_2|A\overline{B}) = 0.69$) who held, a priori, a weaker belief in B, given A, and thus is taken to be more reliable. ■

Conclusion and Outlook

In this paper, we showed how the expressive framework of probability theory and the powerful information-theoretical methodology of optimum entropy (ME-methodology) can be combined to bring forth sophisticated techniques for advanced probabilistic knowledge management. In particular, we presented new approaches to probabilistic belief revision and information fusion by offering solutions to two problems that are usually associated with ME-reasoning: We showed that it is indeed possible to revise certain knowledge by ME-techniques, and we eliminated the weakening effects observed in a naïve ME-fusion approach. This paper continues previous work (Rödder & Xu 1999; Kern-Isberner 2001; Rödder 2000; Kern-Isberner 2002), and once again points out the versatility of the ME-approach.

The generalization of our new ME-revision operator to be applicable for iterated revision, as well as the investigation of the formal properties of the introduced ME-fusion method, will be topics of forthcoming papers.

References

Alchourrón, C.; Gärdenfors, P.; and Makinson, P. 1985. On the logic of theory change: Partial meet contraction and revision functions. *J. Symbolic Logic* 50(2):510–530.

Bloch, I.; Hunter, A.; et al. 2001. Fusion: General concepts and characteristics. *International Journal of Intelligent Systems* 16:1107–1134.

Csiszár, I. 1975. I-divergence geometry of probability distributions and minimization problems. *Ann. Prob.* 3:146–158.

Dubois, D., and Prade, H. 1997. Focusing vs. belief revision: A fundamental distinction when dealing with generic knowledge. In *Proceedings ECSQARU-FAPR'97*, 96–107. Berlin Heidelberg New York: Springer.

Gärdenfors, P. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge, Mass.: MIT Press.

Jaynes, E. 1983. Where do we stand on maximum entropy? In *Papers on Probability, Statistics and Statistical Physics*. Dordrecht, Holland: D. Reidel Publishing Company. 210–314.

Katsuno, H., and Mendelzon, A. 1991. On the difference between updating a knowledge base and revising it. In *Proceedings KR'91*, 387–394. San Mateo, Ca.: Morgan Kaufmann.

Kern-Isberner, G. 1998. Characterizing the principle of minimum cross-entropy within a conditional-logical framework. *Artificial Intelligence* 98:169–208.

Kern-Isberner, G. 2001. *Conditionals in nonmonotonic reasoning and belief revision*. Springer, Lecture Notes in Artificial Intelligence LNAI 2087.

Kern-Isberner, G. 2002. Handling conditionals adequately in uncertain reasoning and belief revision. *Journal of Applied Non-Classical Logics* 12(2):215–237.

Paris, J. 1994. *The uncertain reasoner's companion – A mathematical perspective*. Cambridge University Press.

Reucher, E., and Rödder, W. 2002. Wissensrevision in einer MaxEnt/MinREnt-Umgebung. (manuscript).

Rödder, W., and Kern-Isberner, G. 2003. From information to probability: an axiomatic approach. *International Journal of Intelligent Systems*. (to appear).

Rödder, W., and Meyer, C.-H. 1996. Coherent knowledge processing at maximum entropy by SPIRIT. In Horvitz, E., and Jensen, F., eds., *Proceedings 12th Conference on Uncertainty in Artificial Intelligence*, 470–476. San Francisco, Ca.: Morgan Kaufmann.

Rödder, W., and Xu, L. 1999. Entropy-driven inference and inconsistency. In *Proceedings Artificial Intelligence and Statistics*, 272–277.

Rödder, W. 2000. Conditional logic and the principle of entropy. *Artificial Intelligence* 117:83–106.

Shore, J., and Johnson, R. 1980. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory* IT-26:26–37.