Probabilistic belief expansion and conditioning

Frans Voorbraak*
Department of Mathematics, Computer Science,
Physics and Astronomy
University of Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam, The Netherlands
fransv@fwi.uva.nl

Abstract

The paper discusses probabilistic belief change, in particular the type of change occurring when information is added which is consistent with the old beliefs. In the AGM theory of belief revision this type of belief change is known as an expansion, and is usually considered to be easy compared to revision and contraction, the two other main types of belief changes the AGM theory distinguishes.

If logical theories are used to represent the belief states, then an explicit definition of expansion can be given, and, at first sight, probabilistic conditioning seems a natural generalization of this notion in the context of probabilistic belief states. However, we argue that it may be more correct to view conditioning and expansion as two different kinds of belief change, and we discuss the differences between the two kinds in the context of belief states represented by partially determined probability functions.

1 Introduction

The AGM theory of belief revision [7, 8] has become an important paradigm for investigating rational belief changes. In this theory, three main types of belief changes are distinguished, namely expansion, revision, and contraction. Expansion is the most simple type of belief change, partly because it is supposed to occur only when information is added which is consistent with the previously

*The investigations were carried out as part of the PIONIER-project Reasoning with Uncertainty, subsidized by the Netherlands Organization of Scientific Research (NWO), under grant pgs-22-262.

held beliefs, whereas the other types of belief changes (also) apply in case the new information is inconsistent with the old beliefs.

In fact, if logical theories are used to represent the belief states, then an explicit definition of expansion can be given: The result of expanding a theory K with a sentence ϕ is the set $Cn(K \cup \{\phi\})$ of logical consequences of $K \cup \{\phi\}$. In general, no such explicit definition can be given for revision and contraction.

When these types of belief changes are studied in the context of probabilistic belief states, at first sight, conditioning seems to be the obvious probabilistic variant of expansion. We argue that closer inspection shows that conditioning and expansion are best viewed as two essentially different types of belief changes. This is most clear in the context of partial probability theory, where both types of belief changes can be compared.

In our view, conditioning is a type of belief change different from expansion, revision, and contraction. Conditioning does not make sense in the context of belief states represented by logical theories, just as expansion does not make sense in the context of belief states represented by probability functions. Expansion and conditioning both make sense in the context of belief states represented by partially determined probability functions, since they allow the representation of both ignorance and uncertainty.

In the remainder of this paper, we first review several models of belief states, including belief sets and (partial) probabilistic models. Next, we discuss the notions of expansion and conditioning in the different contexts, and we provide several arguments for our opinion that these notions are essentially different. We conclude with some brief remarks on the relevance of the results for the study of probabilistic versions of revision and contraction.

2 Belief State Models

Throughout the paper, L denotes a propositional language with the usual connectives $\neg$, $\vee$, $\wedge$, $\rightarrow$, $\leftrightarrow$, and constants $\perp$ and $\top$. For simplicity, we assume that L has finitely many propositional letters $p_1, p_2, \ldots, p_n$. We write SL for the set of sentences of L and we use ϕ, ψ, ... as sentential variables.

We further assume $\vdash$ to denote the standard propositional consequence relation on SL and Cn to denote the associated consequence operation. (If $K \subseteq SL$, then $Cn(K) = \{\phi \in SL : K \vdash \phi\}$.) A set $K \subseteq SL$ is called *logically closed* iff $K = Cn(K)$.

Definition 2.1 (belief set) *A* belief set *K for L is a logically closed consistent subset of SL. A sentence ϕ of L is called accepted in K iff $\phi \in K$.*

For technical convenience, the inconsistent belief state (containing all sentences of the language) is sometimes added to the considerations. However, in this paper, the rational belief states are assumed to be consistent.

A sentence $\alpha \in SL$ is called an *atom* of L iff $\alpha = \pi_1 \wedge \pi_2 \wedge \ldots \wedge \pi_n$, where for each $i \in \{1, \ldots, n\}$, $\pi_i = p_i$ or $\pi_i = \neg p_i$. Notice that if n is the number of propositional letters of L, then the number N of atoms of L equals 2^n. A belief set K for L can be represented equivalently by the set of all atoms of L which are consistent with K.

If we define a possible world for L to be an assignment of truth values to all sentences of L according to the usual semantic clauses, then an atom α of L can be identified with a possible world for L, namely the possible world w_α in which α is true. Thus, instead of belief sets one can use sets of possible worlds.

Definition 2.2 (possible worlds model) *A* possible worlds model W *for L is a non-empty set of possible worlds for L. A sentence ϕ of L is called accepted in W iff ϕ is true in every world w of W.*

The belief sets and the possible worlds models are essentially equivalent, but there exist many other essentially different models of belief states. For example, probabilistic models can be used to represent *degrees* of belief in propositions, as opposed to belief sets which only model whether a proposition is (fully) accepted or not.

Definition 2.3 (probabilistic model) *A* probabilistic (belief) model P *for L is a probability function on SL, that is, P is a function on SL satisfying the following three conditions.*

1. *For all $\phi \in SL$, $P(\phi) \geq 0$*
2. *For all $\phi \in SL$, if $\vdash \phi$, then $P(\phi) = 1$*
3. *For all $\phi, \psi \in SL$, if $\vdash \neg(\phi \wedge \psi)$, then $P(\phi \vee \psi) = P(\phi) + P(\psi)$.*

A sentence ϕ of L is called accepted in P iff $P(\phi) = 1$.

We write $PROB(SL)$ for the class of probability functions on SL. Of course, instead of probability functions defined on the sentences of L, one can use probability functions defined on sets of possible worlds for L. We skip the formal definition of this variant.

With every probabilistic model P for L one can naturally associate the belief set $t(P)$ for L given by $t(P) = \{\phi \in SL : P(\phi) = 1\}$. It is easy to see that $t(P)$ is indeed a belief set and that in $t(P)$ the same sentences are accepted as in P. The belief set $t(P)$ is called the *top* of P. Of course, different probability functions may have the same top. The belief sets for L correspond exactly to the equivalence classes of probabilistic models for L with the same top.

In other words, with every belief set K for L one can naturally associate a class of probability functions. Since also a probabilistic belief model P for L can be naturally associated with a class of probability functions, namely $\{P\}$, it follows that using classes of probability functions as models of belief states naturally incorporates the previously defined models.

Definition 2.4 (partial probabilistic model) *A* partial probabilistic model Π *for* L *is a non-empty class of probability functions on* SL. *A sentence* ϕ *of* L *is called accepted in* Π *iff for every* $P \in \Pi$, $P(\phi) = 1$.

Let i be the function embedding belief sets and probability models into partial probability models in the way mentioned above. That is, i assigns to a belief set K for L, or probability function P on SL, the class of probability functions on SL compatible with K, or P. Thus, $i(K) = \{P \in PROB(SL) :$ if $\phi \in K$, then $P(\phi) = 1\}$, and $i(P) = \{P\}$.

The function t assigns belief sets to partial probabilistic models by taking the *top* , i.e., the set of sentences with probability 1. Formally, if Π is a class of probability functions on SL, then $t(\Pi) = \{\phi \in SL : \text{for every } P \in \Pi,\ P(\phi) = 1\}$. The domain of the function t can be extended in a natural way to include belief sets and probabilistic models by defining $t(X) = t(i(X))$ whenever X is a belief set or probability function. It follows that $t(K) = t(i(K)) = K$, whenever K is a belief set, and $t(P) = t(i(P)) = \{\phi \in SL : P(\phi) = 1\}$, whenever P is a probability function.

Partial probabilistic models are of technical interest, since they generalize both the belief sets and the probabilistic models. In addition, it can be argued that in some situations, in particular when there is ignorance concerning the exact likelihoods of events, a class of probability functions is more appropriate to model a belief state than a single probability function. See, for example, [18, 19]. Classes of probability functions are also mentioned in [7] as possible belief state models.

Given a partial probabilistic model Π for L, it is natural to define the *lower probability* of $\phi \in SL$ as $\Pi_{low}(\phi) = \inf_{P\in\Pi} P(\phi)$, and the *upper probability* of $\phi \in SL$ as $\Pi_{up}(\phi) = \sup_{P\in\Pi} P(\phi)$. The belief function of the Dempster-Shafer theory of evidence (see [14]) is a special case of the measure Π_{low} (see [11]). Hence partial probabilistic models also generalize the representation of belief states by such belief functions.

We conclude this section by pointing out that a partial probabilistic model can be viewed as some kind of possible worlds model for a probabilistic language. Let L_{PROB} be a language for probabilistic reasoning with L as object language for the probability expressions. We use SL_{PROB} to refer to the sentences of L_{PROB}, and $\vdash_{PROB}$ and Cn_{PROB} denote a probabilistic consequence relation and associated consequence operation on SL_{PROB}. We assume that the sentences of L are not sentences of L_{PROB}, but occur only in the scope of the probability operator of L_{PROB}.

For example, one can take L_{PROB} to be (a suitable adaptation of) $L(AX)$ or $L(AX_{FO})$ of [5], and use AX_{MEAS} or $AX_{FO-MEAS}$ of the same paper as the probabilistic logic. In [5], the notation w is used instead of the P since the authors also consider logics for inner measures and belief functions. We do not consider such alternative interpretations of the measure, so we can safely use the P.

Sentences of L_{PROB} can be viewed as constraints on probability functions on SL, much as sentences of L can be viewed as constraints on possible worlds for L. Thus, any set $S \subseteq SL_{PROB}$ determines a partial probabilistic model for L. Of course, S and $Cn_{PROB}(S)$ determine the same model. Hence there is an exact correspondence between (probabilistic) belief sets for L_{PROB} and classes of probability functions determined by subsets of SL_{PROB}. For any such class Π, let $\langle \Pi \rangle$ denote the belief set S for L_{PROB} such that $\Pi = \{P \in PROB(SL) : P \text{ satisfies } S\}$. We will sometimes use $P \models S$ as an abbreviation for 'P satisfies S'.

Although, in general, a class of probability functions cannot be expected to correspond to a (probabilistic) belief set for L_{PROB}, it seems likely that any natural occurring class is determined by a set of sentences of some probabilistic language. Therefore, a probabilistic belief set may be an interesting representation of a belief state.

Definition 2.5 (probabilistic belief set) *A* probabilistic belief set *S is a belief set for L_{PROB}, i.e., a logically closed consistent subset of SL_{PROB}. A sentence ϕ of SL_{PROB} is called accepted in S iff $\phi \in S$.*

3 Expansion and Conditioning

Let K be a belief set for L and let $\phi \in SL$ be consistent with K, or, in other words, $\neg\phi$ not accepted in K. Then the belief in ϕ can be added to the belief state K without giving up any old beliefs. In the AGM theory, such a kind of change in belief state is called an expansion, and is modelled by the operation $+$ given by $K + \phi = Cn(K \cup \{\phi\})$.

If we use a probability function P on SL as our representation of belief state, then (Bayesian) conditioning is a natural candidate for the role played by the operation $+$ on belief sets.

Definition 3.1 (Bayesian conditioning) *Let P be a probability function on SL. Define, for each $\phi \in SL$ such that $P(\phi) \neq 0$, the function P_ϕ as follows.*

$$\textit{For every } \psi \in SL, P_\phi(\psi) = P(\psi|\phi) = \frac{P(\phi \wedge \psi)}{P(\phi)}.$$

It is easy to check that the thus defined function P_ϕ is a probability function on SL. The mapping from P to P_ϕ is strongly related to expansion with ϕ, since, just as the expansion operation, the mapping is defined whenever $\neg\phi$ is not accepted in the original belief state ($P(\neg\phi) \neq 1$), and we have $t(P) + \phi = t(P_\phi)$. It follows that the expansion operation $+$ on belief sets can be viewed as an abstraction of Bayesian conditioning on probabilistic belief states.

Of course, the expansion operation on belief sets can be viewed as an abstraction of many other operations on probabilistic belief models as well, including preservative imaging introduced in [7]. The question arises whether compatibility with $+$ is sufficient for an operation on probabilistic belief models to qualify as a probabilistic expansion. This question is perhaps even more relevant in the context of partial probabilistic belief models, where even more operations are compatible with the expansion of belief sets. The following definitions describe two such operations.

Definition 3.2 (extended Bayesian conditioning) *Let Π be a partial probabilistic model for L, and let $\phi \in SL$, such that for some $P \in \Pi$, $P(\phi) > 0$. Define the partial probabilistic model Π_ϕ for L as follows.*

$$\Pi_\phi = \{P_\phi : P \in \Pi, P(\phi) > 0\}.$$

Definition 3.3 (constraining) *Assume that Π is a partial probabilistic model for L and that $\phi \in SL$, such that for some $P \in \Pi$, $P(\phi) = 1$. Define the partial probabilistic model $\Pi_{\&\phi}$ for L as follows.*

$$\Pi_{\&\phi} = \{P \in \Pi : P(\phi) = 1\}.$$

Notice that Π_ϕ is defined whenever $\Pi_{low}(\phi) > 0$, or, in other words, $\neg\phi$ is not accepted in Π, but $\Pi_{\&\phi}$ is only defined when in Π there is no trace of evidence against the possible truth of ϕ, which implies $\Pi_{up}(\phi) = 1$. In the context of belief sets, it may be safe to say that ϕ is consistent with a belief state iff $\neg\phi$ is not accepted in the belief state, but this equivalence is less clear and innocuous in the context of partial probabilistic belief models.

Extended (Bayesian) conditioning is usually called generalized (Bayesian) conditioning, but since we will also consider several other generalizations, we prefer a more informative name. Extended conditioning is essentially a set-extension or component-wise class extension of the ordinary conditioning operation. In the literature, extended conditioning is sometimes only defined for formulas ϕ such that for all $P \in \Pi$, $P(\phi) > 0$. In that case, the definition can be simplified to $\Pi_\phi = \{P_\phi : P \in \Pi\}$, and it is easy to show that extended conditioning is compatible with $+$.

Proposition 3.1 *Let Π be a partial probabilistic belief model of L, and let $\phi \in SL$ such that for all $P \in \Pi$, $P(\phi) > 0$. Then $t(\Pi_\phi) = t(\Pi) + \phi$.*

Proof. $t(\Pi_\phi) = \bigcap_{P\in\Pi} t(P_\phi) = \bigcap_{P\in\Pi}(t(P) + \phi) = \bigcap_{P\in\Pi} Cn(t(P) \cup \{\phi\}) = Cn((\bigcap_{P\in\Pi} t(P)) \cup \{\phi\}) = Cn(t(\Pi) \cup \{\phi\}) = t(\Pi) + \phi$. ∎

However, this restricted version of extended conditioning only allows conditioning on an event ϕ of which one has some evidence in the sense that the lower probability of ϕ is positive. In particular, this means that the natural embedding $i(K)$ of a belief set K in the context of partial probabilistic models can only be (trivially) conditioned with formulas ϕ which are already accepted in K. Therefore, we prefer the less restricted version of extended conditioning, where only the upper probability of the conditioning event is required to be positive.

The following proposition shows that extended (Bayesian) conditioning and constraining are both compatible with $+$ on embeddings of belief sets.

Proposition 3.2 *Let Π be a partial probabilistic belief model for L such that $\Pi = i(K)$, for some belief set K for L, and let $\phi \in SL$ such that $\neg\phi \notin K$. Then $t(\Pi_{\&\phi}) = t(\Pi_\phi) = t(\Pi) + \phi$.*

Proof. Let $\Pi' = \{P \in \Pi : P(\phi) > 0\}$. Similar to the proof of proposition 3.1, we can derive $t(\Pi_\phi) = Cn(t(\Pi') \cup \{\phi\})$. Since $t(\Pi) \subseteq t(\Pi')$ and $\psi \in t(\Pi') \Rightarrow \psi \vee \neg\phi \in t(\Pi)$, we have $Cn(t(\Pi') \cup \{\phi\}) = Cn(t(\Pi) \cup \{\phi\}) = t(\Pi) + \phi$.

It remains to show that $t(\Pi_{\&\phi}) = t(\Pi) + \phi$. $\Pi_{\&\phi} = \{P \in \Pi : P(\phi) = 1\} = \{P \in i(K) : P(\phi) = 1\} = \{P \in PROB(SL) : P(\phi) = 1, \text{and for all } \psi \in K, P(\psi) = 1\} = \{P \in PROB(SL) : \text{for all } \psi \in K \cup \{\phi\}, P(\psi) = 1\} = \{P \in PROB(SL) : \text{for all } \psi \in Cn(K \cup \{\phi\}), P(\psi) = 1\} = i(K + \phi)$. It follows that $t(\Pi_{\&\phi}) = t(i(K + \phi)) = K + \phi = t(i(K)) + \phi = t(\Pi) + \phi$. ∎

Although both operations are compatible with expansions of belief sets, we consider constraining to be the proper probabilistic notion of expansion, whereas (extended) conditioning is in our opinion best viewed as a new kind of operation, different from both expansion and revision. In the rest of this paper we give several arguments supporting this view. The first argument is that, in contrast to conditioning, constraining can be viewed as expansion of probabilistic belief sets.

First notice that constraining can also be defined for sentences of L_{PROB} instead of L.

Definition 3.4 (generalized constraining) *Let Π be a non-empty class of probability functions on SL. Define, for each $\phi \in SL_{PROB}$ which is satisfied by some $P \in \Pi$,*

$$\Pi_{\&\phi} = \{P \in \Pi : P \models \phi\}.$$

This definition of constraining generalizes definition 3.3, since $\Pi_{\&\phi}$, for $\phi \in SL$, can be viewed as an abbreviation of $\Pi_{\&(P(\phi)=1)}$. A discussion of similar generalizations of the notion of conditioning is postponed to the next section. The following proposition shows that the generalized version of constraining translates into the expansion of probabilistic belief sets.

Proposition 3.3 *Assume that S is a probabilistic belief set for L_{PROB}. Let Π be the partial probabilistic belief model for L such that $\langle\Pi\rangle = S$, and let $\phi \in SL_{PROB}$ be consistent with S. Then $\langle\Pi_{\&\phi}\rangle = Cn_{PROB}(S \cup \{\phi\})$.*

Proof. $\Pi_{\&\phi} = \{P \in \Pi : P \models \phi\} = \{P \in PROB(SL) : P \models S \cup \{\phi\}\} = \{P \in PROB(SL) : P \models Cn_{PROB}(S \cup \{\phi\})\}$. It follows that $Cn_{PROB}(S \cup \{\phi\}) = \langle \Pi_{\&\phi} \rangle$. ∎

Corollary 3.1 *Let Π be a partial probabilistic belief model for L such that $\langle \Pi \rangle = S$, and let $\phi \in SL$ such that for some $P \in \Pi$, $P(\phi) = 1$. Then $\langle \Pi_{\&\phi} \rangle = Cn_{PROB}(S \cup \{P(\phi) = 1\})$.*

The following example shows that the analogue of this corollary for conditioning does not hold, provided Cn_{PROB} is monotone, and L_{PROB} is nontrivial, in the sense that it allows assigning (probability) values between 0 and 1.

Example 3.1 *Let p and q be the proposition letters of L, and let Π be the partial probabilistic belief model for L such that $\langle \Pi \rangle = Cn_{PROB}(\{P(p \wedge q) = x, P(p \wedge \neg q) = 0\})$, for some x between 0 and 1. Then $\langle \Pi \rangle$ contains $P(p) = x$, whereas $\langle \Pi_q \rangle$ does not.*

Conditioning is not just a matter of adding information and possibly sharpening the bounds on probability values. Conditioning may also involve the revision of previously held degrees of belief, even if one conditions on events which are completely consistent with the old belief state. The possibility that conditioning is perhaps not a 'pure' expansion process since it has some aspects of a revision process has already been mentioned in [3].

We claim that constraining and conditioning are two essentially different operations. Constraining can be said to reduce ignorance, whereas conditioning is connected with reducing uncertainty. (This will be made more precise later.) Belief sets represent ignorance, but they do not represent uncertainty, whereas probabilistic belief models represent uncertainty, but not ignorance. Only in partial probabilistic belief models both ignorance and uncertainty are represented.

A 'pure' probabilistic expansion operation should reduce ignorance rather than uncertainty, just as the expansion operation $+$ on belief sets. Therefore, such an operation cannot be found in the context of probabilistic belief models. In our opinion, conditioning should not be viewed as a probabilistic version of the expansion operation, but as a new kind of belief change, which does not make sense in the context of belief sets.

4 Expanding from Ignorance

Any belief set K for L can be obtained from the ignorant belief state $Cn(\emptyset)$ by a sequence of expansions. (In fact, since L has a finite number of proposition letters, a single expansion suffices.) This is in accordance with the intuition that one can learn about a subject of which one is completely ignorant, without having to give up previously held beliefs.

Intuitively, one should be able to learn about a subject in bits and pieces, possibly getting information from different sources and on different occasions. Indeed, starting from the ignorant belief state $Cn(\emptyset)$, iterated expansion leads to more and more extended belief states. This is a monotone process, since the following preservation property holds. Let K be a belief set for L, and let $\psi, \phi \in SL$ such that K, ϕ, and ψ are jointly consistent, then

$$\phi \text{ is accepted in } (K + \phi) + \psi. \tag{1}$$

Since for every probability function P on SL and $\phi, \psi \in SL$, such that $P(\phi \wedge \psi) > 0$, we have $(P_\phi)_\psi(\phi) = 1$, the analogous property holds in the context of probabilistic belief models. Here the most obvious candidate to represent the ignorant belief state is the uniform probability function P_{un}, defined by $P_{un}(\alpha) = \frac{1}{N}$, where N is the number of atoms of the language, and α is any one of these atoms. The following example shows that it is *not* the case that any probability function P on SL can be obtained by conditioning the uniform probability function P_{un} on SL.

Example 4.1 *Let p be the only proposition letter of L, and let P be the probability function on SL given by $P(p) = 0.1$. Then P differs from $(P_{un})_p$, $(P_{un})_{\neg p}$, and $(P_{un})_{p \vee \neg p} (= P_{un})$, which are the only probability functions that can be obtained by conditioning P_{un}.*

This negative result can be circumvented if one generalizes the notion of (Bayesian) conditioning by allowing conditioning on events which are not certain. Jeffrey conditioning is such a generalization of Bayesian conditioning. Below we first define a simple version of Jeffrey conditioning, which we call binary Jeffrey conditioning, for reasons that will become clear further on.

Definition 4.1 (binary Jeffrey conditioning) *Let P be a probability function on SL, and let $\phi \in SL$. Define, for any $x \in [0,1]$, the function $P_{\phi,x}$ as follows.*

$$P_{\phi,x} = \begin{cases} xP_\phi + (1-x)P_{\neg\phi} & \text{if } 0 < P(\phi) < 1 \\ P & \text{if } x = P(\phi) = 0 \text{ or } x = P(\phi) = 1 \end{cases}$$

Notice that the usual conditional probability function $P_\phi = P_{\phi,1}$, and that the function $P_{\phi,x}$ is undefined if $P(\phi)$ is either 0 or 1, and $x \neq P(\phi)$.

Example 4.2 *Let p be the only proposition letter of L, and let P be the probability function on SL given by $P(p) = 0.1$. Then $P = (P_{un})_{p,0.1}$.*

Binary Jeffrey conditioning can be generalized in a natural way as follows.

Definition 4.2 (Jeffrey conditioning) *Let P be a probability function on SL, and let $\{\phi_i : i \in I\}$ be a set of mutually exclusive sentences of L, such that for every $i \in I$, $P(\phi_i) > 0$. Define, for any set $\{x_i : i \in I\}$, with $x_i \in [0,1]$ and $\sum_{i \in I} x_i = 1$, the function $P_{\{\langle \phi_i, x_i \rangle : i \in I\}}$ as follows.*

$$P_{\{\langle \phi_i, x_i \rangle : i \in I\}} = \sum_{i \in I} x_i P_{\phi_i}.$$

Notice that if $0 < P(\phi) < 1$, then $P_{\phi,x} = P_{\{\langle \phi, x \rangle, \langle \neg\phi, 1-x \rangle\}}$. Hence binary Jeffrey conditioning is Jeffrey conditioning on *two* exclusive events. Bayesian conditioning is Jeffrey conditioning on a single event and might be called unary Jeffrey conditioning. Jeffrey conditioning can be generalized to allow conditioning on constraints expressed by sentences of L_{PROB} as follows.

Definition 4.3 (minimum cross entropy) *Assume that $\alpha_1, \alpha_2, \ldots, \alpha_N$ are the atoms of L. Let P be a probability function on SL, and let $\phi \in SL_{PROB}$. We define P_ϕ, the* minimum cross entropy update *of P with ϕ, to be that probability function P' on SL satisfying ϕ where the function*

$$I(P', P) = \sum_{i=1}^{N} P'(\alpha_i) \log \frac{P'(\alpha_i)}{P(\alpha_i)}$$

is minimal.

The function $P_{un})_\phi$ is the probability function satisfying ϕ with the maximum entropy. (Entropy is defined in the following section.) It follows that if ϕ uniquely determines a probability function P ($P' \models \phi$ iff $P' = P$), then $(P_{un})_\phi = P$.

It is also easy to show that every probability function P on SL can already be obtained from P_{un} by Jeffrey conditioning. In fact, the following proposition shows that one does not have to start from the 'ignorant' P_{un}, since Jeffrey conditioning allows many probabilistic belief states to be changed into an arbitrary probabilistic belief state.

Proposition 4.1 *Let P be a probability function on SL and let $\{\alpha_i : i \in I\}$ be the set of atoms of L such that $P(\alpha_i) > 0$. Assume that P' is a probability function on SL such that for every $i \in I$, $P'(\alpha_i) > 0$. Then $P'_{\{\langle \alpha_i, P(\alpha_i) \rangle : i \in I\}} = P$.*

Proof. According to the well-known total probability rule, for every $\phi \in SL$, $P(\phi) = \sum_{i \in I} P(\alpha_i) P(\phi | \alpha_i)$. Since for every $i \in I$, $P_{\alpha_i} = P'_{\alpha_i}$, it immediately follows that $P = \sum_{i \in I} P(\alpha_i) P'_{\alpha_i} = P'_{\{\langle \alpha_i, P(\alpha_i) \rangle : i \in I\}}$. ∎

Corollary 4.1 *Let P be a probability function on SL such that for every atom α of L, $P(\alpha) > 0$, or equivalently, $t(P) = Cn(\emptyset)$. Then every probability function on SL can be obtained from P by Jeffrey conditioning.*

In particular, P_{un} satisfies the condition on P mentioned in the above corollary, but there are many other probabilistic belief states other than P_{un} which can be changed into arbitrary probabilistic belief states by means of Jeffrey conditioning. If one allows only binary Jeffrey conditioning, then still a result analogous to proposition 4.1 can be derived.

Proposition 4.2 *Let P be a probability function on SL and let $\{\alpha_i : i \in I\}$ be the set of atoms of L such that $P(\alpha_i) > 0$. Assume that P' is a probability function on SL such that for every $i \in I$, $P'(\alpha_i) > 0$. Then P can be obtained from P' by at most $|I|$ applications of binary Jeffrey conditioning.*

Proof. Let $\alpha_1, \alpha_2, \ldots, \alpha_M$ be the atoms α of L such that $P(\alpha) > 0$. Define $P^1 = P'_{\alpha_1 \vee \ldots \vee \alpha_M}$, and for $1 < m \leq M$, $P^m = P^{m-1}_{\alpha_1 \vee \ldots \vee \alpha_{m-1}, x}$, where

$$x = \frac{\frac{P(\alpha_1 \vee \ldots \vee \alpha_{m-1})}{P(\alpha_m)} \cdot \frac{P^{m-1}(\alpha_m)}{P^{m-1}(\alpha_m \vee \ldots \vee \alpha_M)}}{1 + \frac{P(\alpha_1 \vee \ldots \vee \alpha_{m-1})}{P(\alpha_m)} \cdot \frac{P^{m-1}(\alpha_m)}{P^{m-1}(\alpha_m \vee \ldots \vee \alpha_M)}}.$$

Then for every $i, j \leq m$, $P^m(\alpha_i) : P^m(\alpha_j) = P(\alpha_i) : P(\alpha_j)$. It follows that $P^M = P$. ∎

Example 4.3 *Let p and q be the proposition letters of L, and let P be the probability function on SL given by the table below. Let $P' = P_{un}$. Then $P^1 = (P_{un})_{p \vee q}$, $P^2 = P^1_{p \wedge q, \frac{2}{7}}$, and $P^3 = P^2_{p, 0.9} = P$. The different probability functions are described in the table below.*

	$p \wedge q$	$p \wedge \neg q$	$\neg p \wedge q$	$\neg p \wedge \neg q$
P	*0.4*	*0.5*	*0.1*	*0*
$P' = P_{un}$	*0.25*	*0.25*	*0.25*	*0.25*
P^1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	*0*
P^2	$\frac{2}{7}$	$\frac{5}{14}$	$\frac{5}{14}$	*0*
P^3	*0.4*	*0.5*	*0.1*	*0*

The above example illustrates that the mentioned generalized notions of conditioning do not satisfy the generalization of the preservation property (1). If ϕ and ψ are allowed to range over SL_{PROB}, then ϕ is no longer guaranteed to be accepted in $(P_\phi)_\psi$. Consider, for example, $P = P^1$ from example 4.3, $\phi : P(p \wedge q) = \frac{2}{7}$, and $\psi : P(p) = 0.9$. We conclude that conditioning does not allow a preservative change of the ignorant belief state into an arbitrary belief state.

In the context of partial probabilistic belief models for L, the ignorant belief state is represented by $PROB(SL)$. It is easy to see that any partial probabilistic belief model determined by a subset S of SL_{PROB} can be obtained from $PROB(SL)$ by constraining with the sentences of S. In other words, any probabilistic belief set S for L_{PROB} can be obtained by constraining the ignorant belief state $Cn_{PROB}(\emptyset)$. Moreover, constraining satisfies the appropriate generalization of the preservation property (1):

Proposition 4.3 *Let $S = \langle \Pi \rangle$ be a (probabilistic) belief set for L_{PROB}, and let $\psi, \phi \in SL_{PROB}$ such that S, ϕ, and ψ are jointly consistent. Then*

$$\phi \textit{ is accepted in } (\Pi_{\&\phi})_{\&\psi}.$$

Proof. Since $(\Pi_{\&\phi})_{\&\psi} = \{P \in \Pi : P \models \phi \wedge \psi\}$, any $P \in (\Pi_{\&\phi})_{\&\psi}$ satisfies ϕ. ∎

Of course, one cannot constrain ignorance to partial probabilistic belief models which are not determined by a subset S of SL_{PROB}. Since a similar situation arises in the case of expanding belief sets and possible worlds models of a language with infinitely many proposition letters, we do not consider this to be an essential difference between constraining and expansion.

The situation is much worse for conditioning, since very few partial probabilistic belief models can be obtained from $PROB(SL)$ by extended Bayesian conditioning. Given our previous deliberations, it may be natural to consider set-extensions of the discussed generalizations of Bayesian conditioning. For example, for any $\phi \in SL_{PROB}$, one can define $\Pi_\phi = \{P_\phi : P \in \Pi\}$, where P_ϕ is the minimum cross entropy update of P with ϕ. The following example shows that this operation does not satisfy the preservation property.

Example 4.4 *Assume that p and q are the proposition letters of L, and let $\Pi = PROB(SL)$, $\phi : P(p \wedge q) = 0.5$, and $\psi : P(p) = 0.5$. Then Π, ϕ, and ψ are jointly consistent, but it is not the case that ϕ is accepted in $(\Pi_\phi)_\psi$. The following table shows what happens to P_{un} during the updates.*

	$p \wedge q$	$p \wedge \neg q$	$\neg p \wedge q$	$\neg p \wedge \neg q$
P_{un}	*0.25*	*0.25*	*0.25*	*0.25*
$(P_{un})_\phi$	*0.5*	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$((P_{un})_\phi)_\psi$	*0.375*	*0.125*	*0.25*	*0.25*

Since $P_{un} \in PROB(SL)$, we have $((P_{un})_\phi)_\psi \in (\Pi_\phi)_\psi$. Hence we cannot have $(\Pi_\phi)_\psi(p \wedge q) = 0.5$.

Notice that the example shows that the preservation property is not even satisfied by the set-extension of binary Jeffrey conditioning. We conclude that (iterated) expansion has a certain property, namely the possibility of reaching every (definable) belief state from ignorance in a preservative manner, which is also possessed by constraining, but not by conditioning.

5 Uncertainty versus Ignorance

As previously mentioned, we consider constraining to be an operation which involves reducing ignorance, and conditioning to be aimed at reducing uncertainty. In this section we will try to make this more precise. Although in our opinion ignorance and uncertainty are two distinguishable notions, they are not completely disjoint or unrelated.

Example 5.1 *Consider the following experiments.*

1. *Tossing a coin which is known to be two-headed.*
2. *Tossing a coin which is known to be fair.*
3. *Tossing a coin which is not known to be fair.*

Neither uncertainty nor ignorance is present in the first situation. In the second case, the result is highly uncertain. This of course implies that one is ignorant about the result, but there is no ignorance about the probabilities that represent this uncertainty. In the third case, the result is also highly uncertain, but one is also ignorant about the probabilities involved.

Throughout the paper, we use the term ignorance to refer to the kind of ignorance that is present in the third case of the example above. A good way to explain these informal notions of uncertainty and ignorance is to provide measuring functions for them. In the context of probability functions, entropy is a well-established measure of uncertainty.

Definition 5.1 (entropy) *Assume that $\alpha_1, \alpha_2, \ldots, \alpha_N$ are the atoms of L, and that P is a probability function on SL. The* entropy *$H(P)$ of P is defined as follows. (We assume that $x \log x = 0$ if $x = 0$.)*

$$H(P) = -\sum_{i=1}^{N} P(\alpha_i) \log P(\alpha_i).$$

More generally, if we let p_i denote the value of P on the i-th alternative of an m-valued random variable X, then the entropy *$H_P(X)$ of X is defined as follows.*

$$H_P(X) = -\sum_{i=1}^{m} p_i \log p_i.$$

The subscript P will sometimes be deleted when no confusion is likely to occur.

Notice that one's uncertainty is minimal if one is certain about the truth of some atom α_i, i.e., $P(\alpha_i) = 1$. In that case, $H(P) = 0$. One's uncertainty is maximal if each atom is considered equally likely, and the corresponding entropy

$H(P_{un}) = \log N$. Any $\phi \in SL$ naturally induces a two-valued random variable Φ (with alternatives ϕ and $\neg\phi$), and $H(\Phi)$ can be viewed as a measure of the uncertainty concerning the truth of ϕ.

A first attempt to extend entropy to measure the uncertainty represented by *sets* of probability functions might be to define the uncertainty of a set Π of probability functions to be the maximal entropy of the elements of Π. This does not quite work, since $\max\{H(P) : P \in \Pi\}$ does not necessarily exist. A more serious problem is illustrated in the following example.

Example 5.2 *Assume that p is the only proposition letter of L, and let $\Pi = \{P \in PROB(SL) : P(p) = 1$ or $P(p) = 0\}$. Then $\max\{H(P) : P \in \Pi\} = 0$, even though, based on Π, one is very uncertain about the truth of p.*

To solve both problems, we propose not to maximize over the set itself but over its closed convex hull.

Definition 5.2 (closed convex hull) *Fix some order $\alpha_1, \alpha_2, \ldots, \alpha_N$ of the atoms of L. Then any probability function P on SL can be represented uniquely by a point in $\mathbf{R}^N$. We write P^x for the probability function represented by x. The* closed convex hull *Π^c of a set Π of probability functions on SL is defined to be the smallest superset S of Π satisfying the following two conditions.*

1. *(closure) If $\{P^{x_i} : i \in \mathbf{N}\} \subseteq S$, and $\lim_{i\to\infty} x_i = x$, then $P^x \in S$.*
2. *(convexity) If $P, P' \in S$, and $0 \leq \lambda \leq 1$, then $\lambda P + (1-\lambda)P' \in S$.*

Definition 5.3 (uncertainty) *Let Π be a partial probabilistic belief state for L. The* uncertainty *$U(\Pi)$ of Π is defined as follows.*

$$U(\Pi) = \max\{H(P) : P \in \Pi^c\}.$$

Similarly, the uncertainty $U_\Pi(X)$ of a random variable X is defined as follows.

$$U_\Pi(X) = \max\{H_P(X) : P \in \Pi^c\}.$$

Example 5.3 *Assume that p is the only proposition letter of L, and let $\Pi = \{P \in PROB(SL) : P(p) = 1$ or $P(p) = 0\}$. Then $U(\Pi) = \max\{H(P) : P \in PROB(SL)\} = H(P_{un}) = \log 2$.*

Notice that this uncertainty measure generalizes entropy, since $U(i(P)) = U(\{P\}) = H(P)$. Of course, this is not the only possible generalization. For example, in [10] one can find several generaliz ations of entropy proposed in the context of Dempster-Shafer theory. We do not have a proper justification for our choice, but it suits our main purpose, which is to clarify the difference between uncertainty and ignorance. Similar remarks hold for the measure of ignorance that is defined below.

Given a partial probabilistic belief state Π for L, it is natural to consider the difference between the upper and lower probability of a sentence $\phi \in SL$ to be a measure of the ignorance concerning ϕ. Let us define $I_\Pi(\phi) = \Pi_{up}(\phi) - \Pi_{low}(\phi)$. The following proposition lists some properties of the function I_Π. We omit the straightforward proofs.

Proposition 5.1 *Let Π be a partial probabilistic belief state for L.*

1. *For any $\phi \in SL$, $0 \leq I_\Pi(\phi) \leq 1$.*
2. *$I_\Pi(\top) = I_\Pi(\bot) = 0$.*
3. *For any $\phi \in SL$, $I_\Pi(\phi) = I_\Pi(\neg\phi)$.*
4. *For any equivalent pair $\phi, \psi \in SL$, $I_\Pi(\phi) = I_\Pi(\psi)$.*

We propose to define the total ignorance present in Π as some kind of average of the ignorance concerning the individual non-equivalent sentences of L. Any sentence $\phi \in SL$ is equivalent to a disjunction of atoms of L. Let us fix some order $\alpha_1, \alpha_2, \ldots, \alpha_N$ of the atoms of L. For any $A \subseteq \{1, \ldots, N\}$, we write ϕ_A for $\bigvee_{i \in A} \alpha_i$.

Definition 5.4 (ignorance) *Let Π be a partial probabilistic belief state for L. The* ignorance $I(\Pi)$ *of Π is defined as follows.*

$$I(\Pi) = \frac{\sum_{A \subseteq \{1,\ldots,N\}} I_\Pi(\phi_A)}{2^N - 2}.$$

By proposition 5.1, we can replace the numerator by $\sum_{\emptyset \neq A \subset \{1,\ldots,N\}} I_\Pi(\phi_A)$. Thus $I(\Pi)$ is the average of $I_\Pi(\phi)$, where ϕ runs over all non-equivalent non-trivial sentences of L. It follows that $0 \leq I(\Pi) \leq 1$. As might have been expected, the ignorant partial probabilistic belief state is maximally ignorant: $I(PROB(SL) = 1$. The minimally ignorant belief states are those where the probability function is completely determined: $I(\Pi) = 0$ iff $\Pi = \{P\}$, for some probability function P.

It is interesting to note that definition 5.4 also provides a measure of ignorance for belief sets, since we can define $I(K) = I(i(K))$. The following proposition indicates that the resulting measure behaves quite reasonably.

Proposition 5.2 *Let K, K' denote belief sets for L. Define $I(K) = I(i(K))$.*

1. *$I(K) = 1$ iff $K = Cn(\emptyset)$.*
2. *$I(K) = 0$ iff K is maximal, i.e., for every $\phi \in SL$, $\phi \in K$ or $\neg\phi \in K$.*
3. *If $K \subseteq K'$, then $I(K) \geq I(K')$.*
4. *$I(K) = \frac{r(K)}{2^N - 2}$, where $r(K)$ is the number of consistent $K' \supset K$.*

Both constraining and conditioning can be used to change the ignorant belief state $PROB(SL)$, which is maximally ignorant and maximally uncertain, into a belief state which is minimally ignorant and minimally uncertain, since for any atom $\alpha \in SL$, $I(PROB(SL)_{\&\alpha}) = I(PROB(SL)_{\alpha}) = U(PROB(SL)_{\&\alpha}) = U(PROB(SL)_{\alpha}) = 0$.

Hence we cannot claim that constraining is exclusively used for reducing ignorance and conditioning is only used to reduce uncertainty. However, we will argue that some bias does exist. The first result we mention is that constraining, when defined, is guaranteed to (weakly) reduce ignorance.

Proposition 5.3 *Let Π be a partial probabilistic belief state for L, and let $\phi \in SL_{PROB}$ be satisfied by some $P \in \Pi$. Then $I(\Pi_{\&\phi}) \leq I(\Pi)$.*

Proof. Since $\Pi_{\&\phi} \subseteq \Pi$, we have, for every $\psi \in SL$,

$$\Pi_{low}(\psi) \leq (\Pi_{\&\phi})_{low}(\psi) \leq (\Pi_{\&\phi})_{up}(\psi) \leq \Pi_{up}(\psi).$$

It follows that, for every $\psi \in SL$, $I_{\Pi_{\&\phi}}(\psi) \leq I_{\Pi}(\psi)$. Hence, $I(\Pi_{\&\phi}) \leq I(\Pi)$. ∎

A second property of constraining is that it can be used to strictly reduce the ignorance of all belief states which are not minimally ignorant already.

Proposition 5.4 *Let Π be a partial probabilistic belief state for L such that $I(\Pi) > 0$. Then there exist a $\phi \in SL_{PROB}$ such that $I(\Pi_{\&\phi}) < I(\Pi)$.*

Proof. If $I(\Pi) > 0$, then for some $\psi \in SL$, $\Pi_{up}(\psi) > \Pi_{low}(\psi)$. It follows that for some $P \in \Pi$, $P(\psi) = a$, where $\Pi_{low}(\psi) < a \leq \Pi_{up}(\psi)$. In that case, $I(\Pi_{\&(P(\phi)=a)}) < I(\Pi)$. ∎

This second property can also be proved for Jeffrey conditioning or minimum cross entropy update, since conditioning Π on a sentence defining a probability function $P \in \Pi$ results in a minimally ignorant belief state. However, as shown below, the first result (proposition 5.3) is not true for conditioning.

Proposition 5.5 *Assume that L has at least two proposition letters. There exists a partial probabilistic belief state Π for L such that for some $\phi \in SL$, $I(\Pi_{\phi}) > I(\Pi)$.*

Proof. Assume that p and q are the proposition letters of L, and define

$$\Pi = \{P \in PROB(SL) : P(p) = 0.1, P(\neg p \wedge q) = P(\neg p \wedge \neg q) = 0.45\}.$$

Then $I(\Pi_p) = \frac{8}{14} > \frac{0.8}{14} = I(\Pi)$. ∎

Proposition 5.3 can also be proved for uncertainty, since $\Pi_{\&\phi} \subseteq \Pi$ implies $\max\{H(P) : P \in (\Pi_{\&\phi})^c\} \leq \max\{H(P) : P \in \Pi^c\}$. However, the second property fails in the case of uncertainty, since constraining cannot always be used to reduce the uncertainty of a belief state. For example, the belief state $\{P_{un}\}$ is maximally uncertain, but this uncertainty cannot be reduced by constraining. This suggests that a decrease of uncertainty is merely a byproduct of constraining, whereas a decrease of ignorance can be considered to be the principal product.

It is easy to show that uncertainty in a belief state can always be reduced to zero by conditioning.

Proposition 5.6 *Let Π be a partial probabilistic belief state for L such that $U(\Pi) > 0$. Then there exist a $\phi \in SL$ such that $I(\Pi_\phi) < I(\Pi)$.*

Proof. Let α be an atom of L such that $\neg\alpha$ is not accepted in Π. Then Π_α is a minimally uncertain belief state. ∎

Unfortunately, it is not necessarily the case that $U(\Pi_\phi) \leq U(\Pi)$. In fact, the following example shows that $H(P_\phi)$ may be larger than $H(P)$. (This implies $U(\{P\}_\phi) > U(\{P\})$.)

Example 5.4 *Assume that p and q are the proposition letters of L, and let the probability function P on SL be given by the table below. The table also contains a description of the probability function Pp.*

	$p \wedge q$	$p \wedge \neg q$	$\neg p \wedge q$	$\neg p \wedge \neg q$
P	0.001	0.001	0	0.998
P_p	0.5	0.5	0	0

It is easy to check that $H(P_p) > H(P)$.

Still, there is a sense in which conditioning can be said to reduce uncertainty.

Definition 5.5 (conditional entropy) *Let P be a probability function on SL, and let X and Y be random variables. The* conditional entropy *$H_P(X|Y)$ of X given Y is defined as follows.*

$$\begin{aligned} H_P(X|Y) &= -\sum_{X,Y} P(X,Y)\log P(X|Y) \\ &= \sum_Y (-\sum_X P(X|Y)\log P(X|Y))P(Y). \end{aligned}$$

$H(P_Y)$ is equal to $H_P(X|Y)$, where X is a random variable such that all its values are atoms.

The conditional entropy X given Y is a weighted average of the entropy of X for each value of Y, and it can easily be shown that $H_P(X|Y) \leq H_P(X)$. (See, for example, [6].) This result can be extended to the partial case.

Definition 5.6 (conditional uncertainty) *Let Π be a partial probabilistic belief state for L, and let X and Y be random variables. The* conditional uncertainty $U_\Pi(X|Y)$ *of X given Y is defined as follows.*

$$U_\Pi(X|Y) = \max\{H_P(X|Y) : P \in \Pi^c\}.$$

In particular, the conditional uncertainty $U(\Pi_Y)$ *of Π given Y is given by*

$$U(\Pi_Y) = \max\{H(P_Y) : P \in \Pi^c\}.$$

Proposition 5.7 *Let Π be a partial probabilistic belief state for L, and let Φ be a random variable with alternatives $\phi, \neg\phi \in SL$. Then $U(\Pi_\Phi) \leq U(\Pi)$.*

Proof. The result follows immediately from the fact that for every probability function P, $H(P_\Phi) \leq H(P)$. ∎

In other words, learning about the truth or falsity of some event ϕ is expected to decrease uncertainty, although the actual outcome may result in an increase of the uncertainty. (See example 5.4.)

We conclude that conditioning is better suited than constraining for reducing uncertainty, whereas constraining is the best operation for reducing ignorance.

6 Conditioning versus Constraining

So far, we argued that conditioning or constraining are two different kinds of operations on belief states, where constraining is the proper probabilistic notion of expansion, since it is the expansion of probabilistic belief sets, and it can model preservative changes from the ignorant belief state to an arbitrary belief state. A decrease in ignorance is the main effect of constraining, whereas conditioning is primarily aimed at reducing uncertainty.

We have left open the question which of the two operations one should use when receiving information. We will discuss this matter using a well-known example deriving from Smets [15].

Example 6.1 (The three assassins) *Mr. Jones has been murdered by one of the assassins Peter, Paul, and Mary under orders of Big Boss, who has chosen between these three possible killers as follows. He decided between a male and a female killer by means of tossing a fair coin. A male killer was chosen in case the coin landed heads. Otherwise, a female killer was chosen. No information is available on how he decided between the two male assassins in case the coin landed heads.*

Based on the information above, it seems reasonable to say that the possibility of the killer being male and that of the killer being female are equally likely. Now suppose that you learn that at the time of the murder, Peter was at the police *station, where he was questioned about some other crime. So you can rule out Peter as the killer. How should this new evidence be modelled? In particular, is it still equally likely for the killer to be male or female?*

To formalize this example, let L be the language with the three propositional letters p, q, and r, where p : Peter is the killer, q : Paul is the killer, and r : Mary is the killer. Given that exactly one of the assassins murdered Jones, only three atoms remain possible, namely, $\alpha = p \wedge \neg q \wedge \neg r$, $\beta = \neg p \wedge q \wedge \neg r$, and $\gamma = \neg p \wedge \neg q \wedge r$.

Adding the information that a fair coin toss decided the choice between a male and female killer leads to the partial probabilistic belief state $\Pi = \{P \in PROB(SL) : P(\alpha \vee \beta) = 0.5, P(\gamma) = 0.5\}$. This agrees with the interpretation in Dempster-Shafer theory, where the information is encoded in the mass function m given by $m(\alpha \vee \beta) = 0.5, m(\gamma) = 0.5$, which induces a belief function Bel such that $Bel = \Pi_{low}$. Strict Bayesians will opt for the probability function P given by $P(\alpha) = 0.25$, $P(\beta) = 0.25$, and $P(\gamma) = 0.5$, which is the 'least informative' member of Π.

How to take account of the information $\neg p$ that Peter is not the killer? Strict Bayesians use Bayesian conditioning and arrive at $P_{\neg p}$ given by $P_{\neg p}(\beta) = \frac{1}{3}$, and $P_{\neg p}(\gamma) = \frac{2}{3}$, which implies that it is twice as likely for the killer to be female than to be male. However, as argued by Smets [15] and Halpern and Fagin [9], the 'least informative' prior P on which this answer is based makes some (unjustified) assumptions about how the choice between Peter and Paul is made.

Starting from the partial probabilistic belief state Π, one can use both constraining and (extended) conditioning. Constraining Π with $\neg p$ (or $P(\neg p) = 1$) gives $\Pi_{\& \neg p} = \{P \in PROB(SL) : P(\beta) = 0.5,\ P(\gamma) = 0.5\}$, which implies that the possibility of the killer being male and that of the killer being female are still equally likely. This answer completely agrees with the answer given be Dempster's rule of conditioning in Dempster-Shafer theory, and is defended by Smets [15].

Conditioning Π with $\neg p$ gives $\Pi_{\neg p} = \{P \in PROB(SL) : 0 \leq P(\beta) \leq 0.5,\ 0.5 \leq P(\gamma) \leq 1,\ P(\beta \vee \gamma) = 1\}$, which implies that the possibility of the killer being female is at least as likely as that of the killer being male. This answer, which is defended by Halpern and Fagin [9], agrees with our intuition that finding out that Peter has an alibi makes it less likely that the coin landed heads to a degree which equals one's degree of belief that Peter would have been chosen in case of heads. The ignorance concerning $P(p|p \vee q)$ makes it impossible to justify a specific answer.

The above is related to a distinction introduced by Dubois and Prade [4] between focusing, which concerns a change in reference class, and updating, which concerns a change in belief state. In probability theory, focusing is modelled by conditioning, whereas updating does not necessarily require going to the conditional probability function, but may also be achieved by, for example, applying

Lewis imaging. Dubois and Prade discuss extended conditioning applied to (the set of probability functions compatible with) a Dempster-Shafer belief function, and claim that this rule properly models focusing, and is *not* an updating rule, whereas Dempster's rule of conditioning should be considered to be a kind of updating.

In our opinion, extended conditioning does constitute a kind of updating, and it corresponds to the most common probabilistic belief change operation. Most of the probabilistic evidence an agent encounters is similar to the information that Peter is not the killer in the three assassins example, and should be incorporated by (extended) conditioning.

We do not claim that one should never use constraining. For example, constraining should be used if one does not learn $\neg p$ from Peter's alibi, but from a report of an undercover agent saying that Big Boss decided to choose Paul in the event that a male killer had to be chosen. Constraining is applicable in case of (general) information referring to the prior probabilities, and (specific) information about the case at hand should be incorporated using (extended) conditioning.

Similar remarks can be found in Alechina and Smets [1], although they prefer to model the evidence of Peter's alibi by constraining (or Dempster conditioning), whereas we consider Peter's alibi to be specific information which should be modelled by extended conditioning. Although one cannot formally show that a piece of information like Peter's alibi is specific, and not general, we find the following argument quite convincing.

Example 6.2 *Assume that in the context of example 6.1, you received a report of an undercover agent saying that Big Boss, in the event he has to choose a male killer, decides between Peter and Paul by means of a (second) fair coin toss. Further assume that this report is completely reliable, but that you read it after you learned about Peter's alibi.*

The report of the undercover agent tells you that $P(p)/(P(p)+P(q)) = 0.5$. Simply adding this constraint to $\Pi_{\neg p}$ (as previously defined) results in $\{P \in PROB(SL) : P(\gamma) = 1\}$, which is counterintuitive, and adding the constraint to $\Pi_{\&\neg p}$ is not possible at all, since it leads to an inconsistency. However, the constraint from the report is a constraint on the *prior* probabilities, not on the probabilities obtained after incorporating the evidence of Peter's alibi.

Adding the constraint to Π results in $\{P\}$, where P is the given by $P(\alpha) = 0.25$, $P(\beta) = 0.25$, and $P(\gamma) = 0.5$. Conditioning on the evidence of Peter's alibi gives $\{P_{\neg p}\}$, which agrees with the answer given by strict Bayesians in the original example, since the probability function P is chosen by the strict Bayesians even without the information from the undercover agent. Notice, however, that it is not possible to use constraining to incorporate Peter's alibi in the belief state $\{P\}$. This supports our opinion that Peter's alibi is specific information which calls for conditioning and not for constraining.

The following propositions show that the order among conditionings or among constrainings does not matter.

Proposition 6.1 *Let Π be a partial probabilistic model for L, and let $\phi, \psi \in SL$, such that for some $P \in \Pi$, $P(\phi \wedge \psi) > 0$. Then $(\Pi_\phi)_\psi = \Pi_{\phi\wedge\psi} = (\Pi_\psi)_\phi$.*

Proof. $(\Pi_\phi)_\psi = \{P'_\psi : P' \in \Pi_\phi, P'(\psi) > 0\} = \{(P'_\phi)_\psi : P \in \Pi, P_\phi(\psi) > 0, P(\phi) > 0\} = \{P_{\phi\wedge\psi} : P \in \Pi, P(\phi \wedge \psi) > 0\} = \Pi_{\phi\wedge\psi}$. Similarly, $(\Pi_\psi)_\phi = \Pi_{\phi\wedge\psi}$. ∎

Proposition 6.2 *Let Π be a partial probabilistic model for L, and let $\phi, \psi \in SL_{PROB}$, such that $\phi\wedge\psi$ is satisfied by some $P \in \Pi$. Then $(\Pi_{\&\phi})_{\&\psi} = \Pi_{\&(\phi\wedge\psi)} = (\Pi_{\&\psi})_{\&\phi}$.*

Proof. $(\Pi_{\&\phi})_{\&\psi} = \{P \in \Pi_{\&\phi} : P \models \psi\} = \{P \in \Pi : P \models \psi, P \models \phi\} = \{P \in \Pi : P \models \phi \wedge \psi\} = \Pi_{\&(\phi\wedge\psi)}$. Similarly, $(\Pi_{\&\psi})_{\&\phi} = \Pi_{\&(\phi\wedge\psi)}$. ∎

In contrast, $(\Pi_\phi)_{\&\psi} = (\Pi_{\&\psi})_\phi$ is *not valid*, as shown by the analysis of the example above. Constraining should always be performed *before* conditioning, which implies that after conditioning one should not forget about the original belief state, since general constraints should be added to this original belief state. In this sense, we agree with Dubois and Prade that extended conditioning is not a proper updating rule.

Usually, one obtains additional information from observations of some aspects of the particular case at hand. This kind of information tends to be specific. But the following example shows that in some situations it might be reasonable to search for general information.

Example 6.3 *Consider a chief of police having the information described in the original example 6.1 of the three assassins. He may conclude that he should assign at least as many detectives to investigate Mary as to investigate Paul. However, to determine an optimal division of the available detectives, he might try to reduce his ignorance by ordering an undercover agent to acquire information about the choice between Peter and Paul. Of course, the decision depends on the chief's assessment of the cost and the chance of success of such an undercover operation.*

We conclude that the most common kind of probabilistic evidence is specific evidence about the case at hand, and should be incorporated by means of (extended) conditioning. Some evidence may be of a general nature and may reduce the ignorance concerning prior probabilities. Such evidence calls for constraining.

7 Probabilistic Revision

In the AGM theory, revision is considered to be a belief change which agrees with expansion whenever the new information is consistent with the current belief state, but which is also defined in case the new information contradicts some presently held beliefs.

Definition 7.1 (revision) *Let* $\mathbf{K}$ *be the class of belief sets of* L*, and let* SL_{con} *be the set of consistent sentences of* L*. A function* $*$ *from* $\mathbf{K} \times SL_{con}$ *to* $\mathbf{K}$ *is called a* revision operation *iff it satisfies the following four properties, for every* $K \in \mathbf{K}$ *and* $\phi, \psi \in SL_{con}$*. (We write* $K * \phi$ *for* $*(\langle K, \phi \rangle)$*.)*

1. $\phi \in K * \phi$
2. *If* $K \cup \{\phi\}$ *is consistent, then* $K * \phi = K + \phi$
3. *If* ϕ *and* ψ *are equivalent, then* $K * \phi = K * \psi$
4. *If* $(K * \phi) \cup \{\psi\}$ *is consistent, then* $K * (\phi \wedge \psi) = (K * \phi) + \psi$*.*

The mentioned properties do not uniquely determine a revision operation. Gärdenfors [7] mentions several more or less concrete ways to define revision operations, for example, by using an epistemic entrenchment ordering, or by using partial meet contraction functions. Notice that the notion of revision depends on the notion of expansion. In the context of belief sets, expansion is uniquely determined ($K + \phi = Cn(K \cup \{\phi\})$), but different choices for probabilistic expansion lead to different notions of probabilistic revision.

We have argued that constraining is the proper probabilistic version of expansion. Since constraining is expansion of probabilistic belief sets, one can develop notions of probabilistic revision associated with constraining in a way that exactly mirrors the development of revision in the context of ordinary belief sets.

Alternatively, one might consider Dempster conditioning, which generalizes constraining, since the new evidence is not required to be modelled by a constraint which is consistent with the belief state. For example, if in the case of the three assassins, Mary, instead of Peter, has an alibi, Dempster's rule can be used to condition on $\neg r$, whereas constraining is not defined. However, Dempster's rule is not a full-fledged revision rule since it cannot be used to take account of information that ϕ whenever $\neg\phi$ is accepted. Moreover, it is not clear how to generalize Dempster's rule to partial probabilistic belief states which do not correspond to a Dempster Shafer belief function.

In the literature, one can find some elaborate treatments of probabilistic revision which start from the assumption that conditioning is the correct notion of probabilistic expansion. In the light of our criticism of this assumption, it should be not surprise that these treatments strongly deviate from the development of revision for ordinary belief sets.

Independently of the mentioned assumption, the problem how to generalize conditioning to information which contradicts current beliefs does present itself. Let us call such a generalization of conditioning *c-revision*, to distinguish it from revision generalizing expansion.

Definition 7.2 (c-revision) *Let SL_{con} be the set of consistent sentences of L. A function $*_c$ from $PROB(SL) \times SL_{con}$ to $PROB(SL)$ is called a* c-revision operation *iff it satisfies the following four properties, for every $P \in PROB(SL)$ and $\phi, \psi \in SL_{con}$. (We write $P_{*\phi}$ for $*_c(\langle P, \phi \rangle)$.)*

1. $P_{*\phi}(\phi) = 1$
2. *If $P(\phi) > 0$, then $P_{*\phi} = P_\phi$*
3. *If ϕ and ψ are equivalent, then $P_{*\phi} = P_{*\psi}$*
4. *If $(P_{*\phi}(\psi) > 0$, then $P_{*(\phi\wedge\psi)} = (P_{*\phi})_\psi$.*

An interesting proposal found in, for example, [12, 2] is to define c-revision by means of Popper functions, which are conditional probability functions introduced in [13].

Definition 7.3 (Popper function) *A* Popper function *P on SL is a function from $SL \times SL$ to $[0,1]$ which satisfies the following four properties, for every $\phi, \psi, \chi \in SL$. (We write $P(\phi \uparrow \psi)$ for $P(\langle \phi, \psi \rangle)$.)*

1. $P(\phi \uparrow \phi) = 1$
2. *$P(\cdot \uparrow \phi)$ is a probability function on SL*
3. *If ϕ and ψ are equivalent, then $P(\cdot \uparrow \phi) = P(\cdot \uparrow \psi)$*
4. $P(\phi \wedge \psi \uparrow \chi) = P(\phi \uparrow \chi) \cdot P(\psi \uparrow \phi \wedge \chi)$.

Notice that Popper functions allow the expression of nontrivial constraints on the value of $P(\psi \uparrow \phi)$ even if $P(\phi \uparrow \top) = 0$. If $P(\phi \uparrow \top) > 0$, then $P(\psi \uparrow \phi)$ $= P'(\psi|\phi)$, where $P' = P(\cdot \uparrow \top)$. There is a close relationship between Popper functions and c-revision of probability functions, since for any Popper function P on SL and every $\phi \in SL_{con}$, the mapping from $P(\cdot \uparrow \top)$ to $P(\cdot \uparrow \phi)$ is a c-revision operation (see [12]).

One can extend this approach to *sets* of Popper functions, by defining for any set Π of Popper functions on SL and every $\phi \in SL_{con}$, the mapping from $\{P(\cdot \uparrow \top) : P \in \Pi\}$ to $\{P(\cdot \uparrow \phi) : P \in \Pi\}$. However, the use of Popper functions to define c-revision runs into serious difficulties. For example, it is not clear how to treat iterated c-revision using this approach (see [2]). Perhaps even more serious is the problem indicated below.

Example 7.1 *Mr. Jones has been murdered by one of the assassins Peter, Paul, and Mary under orders of Big Boss, who has chosen between these three possible killers as follows. He draws a coin which is either fair of two-tailed. He tosses the chosen coin twice. A female killer is chosen if the first result is tails. Otherwise, a male killer is chosen, where Peter is chosen if the second result is also heads, and Paul is chosen if the second result is tails.*

Now suppose that you learn that at the time of the murder, Mary was at the police station, where she was questioned about some other crime. So you can rule out Mary as the killer. Intuitively, the remaining possibilities, Peter is the killer or Paul is the killer, are equally likely and get probability 0.5.

Since $P(p \uparrow \neg r) = 0.5$ and $P(p \uparrow \neg r) = 0$ are both compatible with the general information, c-revision with the specific information $\neg r$ does not lead to the intuitively correct answer. More generally, the example indicates that one cannot simply define $\Pi_{*\phi} = \{P_{*\phi} : P \in \Pi\}$. If $P(\phi) > 0$, for some $P \in \Pi$, then one should have $\Pi_{*\phi} = \Pi_{\phi} = \{P_{\phi} : P \in \Pi, P(\phi) > 0\}$. Even in case $P(\phi) = 0$, for all $P \in \Pi$, it remains questionable whether it is reasonable to revise using the available conditional beliefs.

Example 7.2 *Consider example 7.1, but now assume that it is given that the two-tailed coin is drawn. Intuitively, it is not at all clear that Paul is the killer. The available information is inconsistent and this inconsistency can be resolved by questioning any piece of general information and any information on which one has previously conditioned.*

A related, and perhaps more convincing example is the following one, which derives from Stalnaker [16].

Example 7.3 *Suppose I accept that if Hitler had decided to invade England in 1940, Germany would have won the war. Then suppose I discover, to my surprise, that Hitler did in fact decide to invade England in 1940 (although he never carried out his plan). In that case, it seems reasonable for me not to conclude that Germany has won the war, but to give up my belief in the conditional.*

Using Popper functions for c-revision gives a special status to conditional information, which perhaps can be defended for hypothetical reasoning, but not for actual belief change. Similar remarks concerning using the Ramsey test for revision are made in [17]. In order to express conditional information one, in general, needs something like Popper functions or counterfactuals, but this conditional information should get no special status for belief changes.

A possible approach to c-revision would be to develop a language in which one can express both general constraints, including those representing counterfactuals, as specific information on which one has previously conditioned. One can then follow the way ordinary notions of revision are defined.

We conclude that probabilistic revision is quite similar to ordinary revision, provided one recognizes constraining as the proper probabilistic version of expansion. As for generalizing conditioning to new information inconsistent with the current beliefs (c-revision), the popular approach using Popper functions is troublesome.

8 Conclusion

We argued that conditioning can best be viewed as a type of belief change different from expansion, revision, and contraction. This difference becomes most clear in the context of belief states represented by partially determined probability functions, which allow the representation of both ignorance and uncertainty. In this context, there are several operations agreeing with expansion on belief sets. Of these, constraining has more right to be called probabilistic expansion, than conditioning has, although the latter is often chosen in the literature.

The principal result of constraining is a decrease in ignorance, whereas conditioning is aimed at reducing uncertainty. General evidence reducing ignorance concerning the prior probabilities calls for constraining, but the most common probabilistic evidence concerns the particular case at hand, and should be modelled by means of conditioning. Constraining should always be applied *before* conditioning, since it represents evidence concerning the *prior* probabilities.

Conditioning does not make sense in the context of belief sets, which do not represent uncertainty, but only ignorance, just as expansion does not make sense in the context of belief states represented by probability functions, which do not represent ignorance, but only uncertainty. Our distinction between expansion and conditioning calls into question the treatments of probabilistic revision which start from the assumption that conditioning is the correct notion of probabilistic expansion.

References

[1] N. Alechina and P. Smets, A note on modal logics for partial belief, *ILLC research report X-94-06*, September 1994.

[2] C. Boutilier, On the revision of probabilistic belief states, *Notre Dame Journal of Formal Logic* 36 (1995) 158–173.

[3] D. Dubois and H. Prade, Belief change and possibility theory, in: P. Gärdenfors, ed., *Belief Revision* (Cambridge U.P., Cambridge, 1992) 142–182.

[4] D. Dubois and H. Prade, Focusing versus updating in belief function theory, in: R.R. Yager, J. Kacprzyk, and M. Fedrizzi, eds., *Advances in the Dempster-Shafer Theory of Evidence* (Wiley, New York 1994) 72–95.

[5] R. Fagin, J.Y. Halpern and N. Megiddo, A logic for reasoning about probabilities, in *Proc. 3rd IEEE Symp. on Logic in Computer Science* (1988) 277–191.

[6] R.G. Gallagher, *Information Theory and Reliable Communication* (Wiley, New York, 1968).

[7] P. Gärdenfors, *Knowledge in Flux. Modeling the Dynamics of Epistemic States* (MIT Press, Cambridge MA, 1988).

[8] P. Gärdenfors, Belief revision: An introduction, in: P. Gärdenfors, ed., *Belief Revision* (Cambridge U.P., Cambridge, 1992) 1–28.

[9] J.Y. Halpern and R. Fagin, Two views of belief: belief as generalized probability and belief as evidence, *Artificial Intelligence* 54 (1992) 275–317.

[10] G.J. Klir, Measures of uncertainty in the Dempster-Shafer theory of evidence, in: R.R. Yager, J. Kacprzyk, and M. Fedrizzi, eds., *Advances in the Dempster-Shafer Theory of Evidence* (Wiley , New York, 1994) 35–49.

[11] H.E. Kyburg, Bayesian and non-Bayesian evidential updating, *Artificial Intelligence* 31 (1987) 271–293.

[12] S. Lindström and W. Rabinowicz, On probabilistic representation of non-probabilistic belief revision, *J. of Philosophical Logic* 18 (1989) 69–101.

[13] K.R. Popper, *The Logic of Scientific Discovery* (Hutchinson, London, 1959).

[14] G. Shafer, *A Mathematical Theory of Evidence* (Princeton U.P., Princeton, 1976).

[15] P. Smets, Belief functions versus probability functions, in: B. Bouchon, L. Aitt, R.R. Yager, eds., *Uncertainty and Intelligent Systems*, LNCS 313 (Springer, Berlin, 1988) 17–24.

[16] R. Stalnaker, *Inquiry*, (MIT Press, Cambridge MA, 1984).

[17] F. Voorbraak, Conditionals, probability, and belief revision, in M. Stokhof and L. Torenvliet, eds., *Proc. 7th Seventh Amsterdam Colloquium* (University of Amsterdam, 1990) 597–613.

[18] F. Voorbraak, *As Far as I Know. Epistemic Logic and Uncertainty*, (Dissertation, Utrecht University, Utrecht, 1993).

[19] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, (Chapman and Hall, London, 1991).