Reliability and Imprecision: FEW 2005 comments on Haenni and Hartmann

Gregory Wheeler

Artificial Intelligence Center - CENTRIA
Department of Computer Science, Universidade Nova de Lisboa
2829-516 Caparica, Portugal
greg@di.fct.unl.pt

1 Introduction: The notion of partial reliability

When we say of something that it is unreliable we are not necessarily saying that the chances of it behaving as one expects equal the chances of it behaving otherwise. The train service in Lisbon is reliable, I tell you, yet the ambulance service in Lisbon is unreliable. By this, I am not saying that I've never boarded a late commuter train in Lisbon, nor am I saying that it is a toss of a coin whether dialing 112 will fetch you an ambulance. Indeed, you would find it surprising were this to be what I meant by 'reliable' and 'unreliable'. Moreover, if this *was* what I believed to be to true of Lisbon trains and ambulances, you'd think I was understating each case by putting it this way. Either way, you would find both the random sense of unreliable and the "with certainty" sense of reliable extreme.

Yet, this is the concept of unreliability used in the Haenni-Hartmann (HH) model. In their introduction of reliability variables, REP_i, indexed to an information source i, Haenni and Hartmann write that

> if source i is unreliable, then we suppose the outcome of REP_i to depend only on a randomization parameter r_i no matter what the true state of [the hypothesis i is reporting on] HYP is. In other words, we assume $Pr(Rep_i|Hyp, \neg Rel_i) = Pr(Rep_i|\neg Hyp, \neg Rel_i) = r_i$ [3, p. 4].

This observation raises the question of how to modify the HH model for partial reliability to effect a more realistic representation of the reliability. This note addresses this question through a few remarks about the mathematical underpinnings of the HH model, namely Dempster-Shafer (DS) theory, and a remark about a particular independence assumption, called MAR, implicit in the HH model. It turns out that MAR is a strong and in many cases unreasonable assumption to make. To weaken this assumption, however, one should be able to distinguish between facts about the reason for receiving incomplete information (called a *protocol* or *incompleteness mechanism*) and information provided solely by the values of an attribute vector. This, in turn, requires a more expressive language than D-S theory can provide. The notion of *lower previsions* from the theory of imprecise probabilities offers an attractive alternative.

However, one of the primary reasons for adopting simplifying assumptions is to facilitate implementation. Indeed, implementation considerations appear to be the strongest reason for preferring D-S theory to the theory of imprecise probabilities, since we have 20 years of experience working with D-S theory in expert systems and the natural graphical representation of lower previsions, credal nets, is much more complex than standard bayes nets. However, recent results of Gert de Cooman and Marco Zaffalon [1] offer an update rule that classifies incomplete evidence, yielding a set of dominating possible optimal classes in linear time (in the size of input) for singly connected subgraph [1, p. 109]. Thus, the availability of these rules reduces the relevant search space withing Credal nets, making the later a tractable representational framework.

2 Protocols

The assumption behind the stochastic interpretation of reliability in the HH model is that there is no systematic reason for a source to be unreliable. We'll return to this assumption in subsection 2.2. For now, it is important to notice that it is a strong assumption to make about the problem domain, one that is not always satisfied. We may illustrate this by considering the Monty Hall puzzle.

Example 1 (Monty Hall puzzle). Imagine game show host Monty Hall directing the attention of a member of his audience to three doors. A new Chevrolet Vega is behind one of these doors, he tells the contestant, and a goat is behind each of the other two. Monty asks the contestant to choose a door. After the contestant makes her choice, Monty opens one of the remaining doors to reveal a goat. He then asks the contestant if she'd like to switch her choice. Assuming the contestant prefers a Vega to a goat, the question is what she should do.

The correct answer depends on the *protocol* [9] that Monty follows in selecting the door. What the contestant knows about this protocol specifies the *incompleteness mechanism* [1] of the situation. Consider the following example protocols, all consistent with what we know about the example:

1. (*Satanic Monty*) Suppose Monty knows where the Vega is but shows a goat *only* when the contestant has picked the Vega; otherwise, he shrugs.
2. (*Give-away-the-store Monty*) Suppose Monty shows a goat *only* when the contestant has picked a goat; otherwise, he shrugs.

The point is, knowing what protocol describes the situation will determine the set of options reasonable for the contestant to consider. Once we realize that the conditioning event—Monty showing her a goat—alone is compatible with all of these protocols. Thus, the corresponding incompleteness mechanisms:

1. If the incompleteness mechanism is described by *Satanic Monty*, then the contestant should with certainty *not* switch: the contestant would know upon seeing a goat that she chose the Vega.
2. If the incompleteness mechanism is described by *Give-away-the-store Monty*, then the contestant should with certainty switch: the contestant would know upon seeing a goat that the Vega is behind the remaining door.

2.1 What should the contestant do?

Answer: complain.

That is, if all that she knows about Monty's protocol is what is given in the initial example, then her seeing a goat is no grounds for her to judge that one door is a better choice than another *nor to judge that one is not better than the other*. In other words, given that Monty's shown her a goat, the probability of selecting the Vega is in the interval $[0, 1]$: conditioning on the event of seeing a goat is, in *this version of the puzzle*, vacuous. The most reasonable thing for her to do is to ask Monty to tell her what he's up to.[1]

2.2 Ignorance, Imprecision and MAR

This analysis is in stark contrast to an assumption standardly made in uncertainty frameworks, including Haenni and Hartmann's model for partial reliability. The assumption is the *missing at random* assumption (MAR), which states that conditioning on missing information about a random variable is the same as conditioning on any possible value of that variable. In other words, MAR assumes that there is no systematic reason for the missing value.[7]

MAR appears in Haenni and Hartmann's model when they first introduce their primitive notion of unreliability in their second example, which we quoted in the introduction.

The important point here is this. MAR is an assumption that holds only in special cases of updating probabilities on incomplete or set-valued information. In so far as we wish to build a general model for updating on incomplete or partial information, then, we should not adopt a representation of uncertainty that builds in this assumption. Instead, we should adopt a framework that allows us to make explicit and transparent all of the uncertainty assessments and assumptions behind the framework. Much like the conditions for good governance include accountability and transparency of decision making, so too should our uncertainty framework be accountable and facilitate transparent decision making by making all uncertainty assessments explicit. In the present context, this includes the ability of a system to represent partial or complete ignorance in addition to partial or conflicting information and imprecise assessments or measurements of uncertainty [15].

Sometimes the most reasonable thing to do is to suspend belief and go back to the problem domain to get more information. Our models of uncertainty should

[1] FEW participants may recall Teddy Seidenfeld's remark apropos of this advice, namely whether the contestant should (in effect) pay Monty not to tell her his choice on the grounds that a probability of $\frac{1}{3}$ is preferable to the unit interval. For a discussion of dilation, see [8] [4]). Teddy's point is that it cannot *always* be the most reasonable thing to do, to pay to have information withheld. However, and I think he agrees on this point as well, there clearly are cases where what we should do is demand more information from the problem domain. There is a notion of uncertainty involved here that is relatively new and not yet understood.

be sophisticated enough to signal to us when this is the appropriate course of action to take.

The problem here is to explain, analytically, why naive conditioning is not a reliable method. The answer lies in specifying what information is learned about the incompleteness mechanism that prohibits a precise assessment of uncertainty. Haenni and Hartmann are quite right to focus on the role that ignorance plays in modeling partial reliability. This is one of the key ideas to Dempster-Shafer (DS) theory. But it is not the complete picture, and it is instructive to consider why not.

2.3 Representing Protocols

An incompleteness mechanism transforms a complete (or point valued) observation into an incomplete (or interval valued) one. Neglecting the incompleteness mechanism, when known, results in naive applications of conditioning. The Monty Hall puzzle is an example of problematic results of naive conditioning.

Representing the incompleteness mechanism of a state is a key component to modeling reasoning about partially reliable information sources. Following [14] and recent results of [1], we may do this in terms of the theory of *coherent lower previsions*. It turns out that Demster-Shafer theory is a special case of the theory of lower previsions.

3 Dempster-Shafer and Imprecise Probabilities

Dempster-Shafer (DS) theory may be considered as consisting of two components. First, there is the notion of a *belief function*, written $Bel(\cdot)$, which is a real-valued function defined on subsets of a sample space (frame of discernment), Ω, that is

$$Bel(A) = \sum_{B \subseteq A} m(B),$$

for all $A \subseteq \Omega$, where m is a function on subsets of Ω satisfying

(m1) $m(\emptyset) = 0$;[2]
(m2) $m(B) \geq 0$, for all $B \subseteq \Omega$; and
(m3) $\sum_{B \subseteq \Omega} m(B) = 1$.

The mass function m is a *probability assignment* for $Bel(\cdot)$. The mass function m is determined by $Bel(\cdot)$ through the Möbius inversion formula

$$m(B) = \sum_{A \subseteq B} (-1)^{|B-A|} Bel(A).$$

Observe:

[2] Note that Smets and Kennes' [12] presents a version in which $m(\emptyset) \geq 0$.

- All *Bayesian measures* are belief functions but note vice versa. A bayesian measure is the limiting case when $Bel(B) = 0$ for all $B \subseteq \Omega$ unless B is a singleton set.
- The function Bel is a *lower probability*, modulo interpretation of the function. The key point is that the measure is defined on *events* rather than *gambles* and, hence, is a less expressive representational language than the language of previsions. (More below).
- There are some types of incomplete information that Bel functions cannot represent properly. See (Wally 1996, p. 28) or (Halpern 2003, Ch 2) for examples and discussion.

The second component is *Dempster's Rule of updating*, which is a two place function $m \oplus m'$ taking mass functions m and m' as arguments and returning a new "combined" mass function, m''. The idea behind Dempster's rule is to view discrete mass functions, m_1 and m_2, as representing "independent" sources of information. The idea is to suppose that observation O_1 comes from one source while O_2 comes from another source of information. The evidence for the combined observations O_1 and O_2, on this view, should consist of all the reports for each observation such that $O_1 \cap O_2$.

However, there is a strong *conditional* independence assumption buried here that is often confused for a weaker (and more plausible) unconditional independence assumption. Put in terms of a reliable witness, the two may be distinguished as follows:

- *Unconditional independence*: Assumes that witnesses are *known* to not have any interaction. Then it is reasonable to assume that they are unconditionally independent.
- *Conditional independence*: holds only if a witness's belief about the correctness of her report would be unchanged by further information, including other witness reports and whether those reports are correct.

Note that Shafer's justification for Demster's rule (Shafer 1981, 1990, 1991) suggests unconditional independence, but that this independence assumption is insufficient; one needs conditional independence. For discussion on precisely this point, see (Levi's [6]).

Dempster's rule doesn't always fail, of course. But the important point here is that the independence assumptions that are integral to the rule are not explicit parameters that we may relax. To do this we need a more expressive framework.

3.1 Relationship between DS and Previsions

It turns out that DS-theory can be viewed as a special case of the theory of imprecise probabilities [14]. Viewed from this perspective, the main distinguishing feature of Dempster-Shafer theory is Dempster's rule of combination, or the second component of the theory. There is a quite straight-forward reading of

Belief functions as Lower Previsions, and corresponding mass functions as lower probabilities. The conflict between the two theories comes over Dempster's rule. But moving to a more expressive framework—working with previsions rather than probabilities—would dispense with the need for Dempster's rule, which is an attempt to get more precise measures of uncertainty by combining events than would otherwise follow. However in some cases, less precise measures are precisely the correct result given the values assigned to constituents *and nothing more.*

3.2 Imprecise probabilities

Let X be a random variable taking values in a finite set $\mathcal{X}$. A gamble f on the value of X is a function, $f : \mathcal{X} \mapsto \mathbb{R}$, where $\mathbb{R}$ is interpreted as an ordered set of rewards. For any value $x \in \mathcal{X}$, $f(x)$ assigns a "value" from $\mathbb{R}$. If a subject is uncertain about what value from $\mathcal{X}$ the random variable X takes, then he will be disposed to accept certain gambles or reject others, and we may model this uncertainty, behaviorally, by observing what gambles he accepts and what gambles he rejects.

The Bayesian picture of uncertainty assumes that a subject may always specify the *fair price* or *prevision*, $P(f)$, for any gamble f, whatever the information available to him. On this picture, $P(f)$ is the unique real number such that the subject:

- accepts $f - p$, accepts to buy the gamble f for a price p, for all $p < P(f)$.
- accepts $q - f$, accepts to sell the gamble f for a price q, for all $q > P(f)$.

A consequence of this view is that, for any real number r, the available information allows a subject to decide which of the following options he prefers: buy f for price r, or sell f for that price. A property of P on this picture is that P is *linear* in the sense that, for all gambles f and g on $\mathcal{X}$ and all non-negative real numbers λ, P satisfies

(P1) $\min_{x \in \mathcal{X}} f(x) \leq P(f)$ (*Accept sure gains*)
(P2) $P(f + g) = P(f) + P(g)$ (*Additivity*)
(P3) $P(\lambda f) = \lambda P(f)$, for any $\lambda \in \mathbb{R}_+$ (*Positive Homogeneity*)

A *coherent linear prevision* is simply a prevision P satisfying $P1$, $P2$ and $P3$. These pick out the same thing as de Finetti's *coherent previsions*, and are sometimes called *precise* probability models.

It has been argued in the literature [13, 5, 14] that the assumption that agents always may satisfy the *fair price assumption* is not warranted; there may be cases where an agent may not be disposed to buy or sell a gamble at a particular price, based on the available information. In such cases we often say that the two outcomes are *not comparable.*

The theory of imprecise probabilities [14, 15] proposes to remedy this by allowing the subject to specify two numbers, $\underline{P}(f)$ and $\overline{P}(f)$. The subject's *lower* *prevision for f* $\underline{P}(f)$ is the greatest real number p at which he will buy gamble f; the subject's *upper prevision for f* $\overline{P}(f)$ is the least real number q at which he will sell gamble f. For any r that is greater than the subject's *lower prevision for f* and less than the subject's *upper prevision for f* is a point whereby the subject does not express a preference for buying or for selling f at r.

Selling a gamble f for price r is the same thing as buying $-f$ for price $-r$, so the following conjugacy relation holds

$$\overline{P}(f) = -\underline{P}(-f) \tag{1}$$

In addition to being able to define lower previsions in terms of upper previsions, we may also see linear previsions as being a special case of *lower* and *upper* previsions. Taking lower previsions to be primitive, for all gambles f and g on $\mathcal{X}$ and all non-negative real numbers λ, $\underline{P}$ satisfies

($\underline{P}1$) $\min_{x\in\mathcal{X}} f(x) \leq \underline{P}(f)$ (*Accept sure gains*)
($\underline{P}2$) $\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g)$ (*Super-additivity*)
($\underline{P}3$) $\underline{P}(\lambda f) = \lambda P(f)$, for any $\lambda \in \mathbb{R}_+$ (*Positive Homogeneity*)

De Finetti's [2] theory then is the special case of a coherent lower prevision that is also *self-conjugate* or *maximally precise*, that is

$$\underline{P}(f) = \overline{P}(f), \text{ for all gambles } f.$$

The difference between *linear* and *lower* previsions is reflected by (P2) and ($\underline{P}2$).

It follows that for all $\underline{P}$ satisfying ($\underline{P}1$), ($\underline{P}2$) and ($\underline{P}3$),

$$\overline{P}(f) \geq \underline{P}(f), \text{ for all } f \in \mathcal{L}(\mathcal{X}). \tag{2}$$

3.3 Lower Probabilitiy and Lower Previsions

The lower/upper probability of an event A can be interpreted as specifying acceptable rates for betting on or against A. One way to construct a lower prevision $\underline{P}$ is to evaluate all upper and lower probabilities and then construct the lower prevision by natural extension. Suppose that $(p_1(X_1), ..., p_n(X_n))$ is a sequence of probability assignments assigned to outcome $X_1, ..., X_n$. A *natural extension* is a set $\mathcal{M}$ of all mass probability mass functions

$$\{(p_1(X_1), ..., p_n(X_n))_1, ..., (p_1(X_1), ..., p_n(X_n))_k\}$$

that are consistent with the set of constraints encoded by prior uncertainty constraints, expressed as a system of linear equalities on the probability mass functions, the p_i's.

It is shown in [14] and remarked on in [15, 1] that lower previsions contain more information than lower probabilities in the sense that not all lower previsions can be constructed from lower probabilities. (Conversely, upper and lower probabilities are uniquely determined by upper/lower previsions.)

The extra information contained in lower previsions is important, particularly when conditioning. Conditioning on upper and lower probabilities may be too imprecise, since one may wind up conditioning on the extreme points defined by the natural extension.

This last point brings us back to considering the representing uncertainty, belief functions and Bayesian networks. The underlying representation of uncertainty to Haenni and Hartmann's framework is Dempster-Shafer belief functions, which is a species of probability theory. Points to notice here:

- Every belief function is a coherent lower probability function.
- Both Belief functions and Lower Previsions *can but need not* be interpreted as a lower bound (envelope) of a (set of) probability measure(s). Shafer is emphatic that belief functions not be tied to a *Bayesian Sensitivity analysis interpretation*. Each is compatible with this interpretation, but is more general than and so does not presuppose it.
- The main difference between Lower Previsions and DS theory is Dempster's rule for combining belief functions. This rule relies on strong independence assumptions that are not always satisfied. On the Lower Previsions view, there is no justification for the *indiscriminate* use of Dempster's rule.

Hence, we may replace D-S theory with Lower Previsions without loss of expressive capacity in the framework.

The remaining motivation for preferring D-S theory to Lower Previsions stems from implementation concerns. There is a well-established literature on using Dempster's rule of combination to combine belief functions on certain kind of trees [11, 10, **?**], Expert systems in 1985, 1988 and 1989, which naturally leads to their use in Bayes Nets. It what remains we highlight recent work by Gert de Cooman and Marco Zaffalon [1] that addresses this point.

4 Conditioning on Incomplete Information

Suppose there is a random variable X, taking values from a finite set of values $\mathcal{X}$ Suppose we have a model for the available information about what value X will assume in $\mathcal{X}$, represented by a coherent lower prevision $\underline{P}_0$ defined on the set of all gambles, $\mathcal{L}(\mathcal{X})$.

Suppose now that additional information about the value of X is obtained by observing the value of another random variable, O, which takes values in a finite set of values $\mathcal{O}$. The information gathered, however, is incomplete in the sense that the value of O does not uniquely determine the value of X. Rather, the only information that we have about the relationship between X and O is that if we know that X takes a value $x \in \mathcal{X}$, then O must take a value o from the non-empty subset $\Gamma(x)$ of $\mathcal{O}$, *and nothing more*.

What, then, do we know about X after observing O? We know that its value must be among those values of X that produce the observation $O = o$, written

$$\{o\}^* = \{x \in \mathcal{X} : \Gamma(x)\}$$

Unless $\{o\}^*$ is a singleton, the observation $O = o$ does not uniquely determine the value for X; rather, it allows us to restrict the possible values of X to the set $\{o\}^*$. This condition also holds even in the case where there is some possible value X such that o is the only compatible observation, that is if

$$\{o\}_* = \{x \in \mathcal{X} : \Gamma(x) = (o)\}$$

is non-empty. The set $\{o\}^*$ includes $\{o\}_*$ but may contain more than one element.

This is the basic setup for the problem of conditioning on incomplete information.

4.1 MAR/CAR

CAR, *Coarsening at random*, is a generalization of MAR, *Missing at random*. It is an assumption invoked when there is an incomplete—or missing, or unreliable—observation report, $O = o$. CAR states that the probability of observing $O = o$ is not affected by the values x of X, that is

$$p(o|x) = p(o|y) > 0,$$

for all $o \in \mathcal{O}$ and all x and y in $\{o\}^*$ such that $p_0(x) > 0$ and $p_0(y) > 0$. Applying Bayes rule, we derive from CAR that

$$p(x|o) = \begin{cases} \frac{p_0(x)}{P_0(\{o\}^*)} = p_0(x|\{o\}^*) & \text{if } x \in \{o\}^*, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

This means that making the CAR assumption about the incompleteness mechanism warrants using the so-called *naive updating rule*, (3).

One result of de Cooman and Zaffalon (2004) is that naive updating is warranted—even if we know nothing of the protocol—when, and only when,

$$\{o\}_* = \{o\}^*,$$

that is, when all the states that *may* produce observation o can *only* produce observation o.

CAR, as has been noted here and in the literature, is a strong assumption to make about the reason that missing information is missing. Specifically, CAR claims that there is no systematic reason for the data to be missing.

Before investigating a weakening of CAR/MAR, we must take a quick detour through a mechanism for representing the information between values for X and O, one that relies on the definition of a multi-valued map.

Vacuous Lower Previsions Model the claim that 'O assumes a value in $\Gamma(x)$' as the *vacuous lower prevision* $\underline{P}_{\Gamma(x)}$ on the set of gambles $\mathcal{L}(\mathcal{O})$ relative to the subset $\Gamma(x)$ of O. The relationship between X and O may then be modeled by the (vacuous) conditional lower prevision $\underline{P}(\cdot|X)$ on $\mathcal{L}(\mathcal{O})$, defined by

$$\underline{P}(g|x) = \underline{P}_{\Gamma(x)}(g) = \min_{o \in \Gamma(x)} g(o) \tag{4}$$

for any gamble g on $\mathcal{O}$.

Note that the following *Bayesian sensitivity analysis* interpretation of equation (4). Let $\mathcal{M}$ be the set of dominating linear previsions such that

$$\mathcal{M}(\underline{P}) = \{P \in \mathcal{P}(\mathcal{X}) : (\forall f \in \mathcal{L}(\mathcal{X}))(\underline{P}(f) \leq P(f)\}.$$

Then the coherent lower prevision $\underline{P}(\cdot|x)$ is the lower envelope of the set

$$\mathcal{M}(\underline{P}(\cdot|x)) = \{P(\cdot|x) : P(\mathit{\Gamma}(x)|x) = 1\}$$

of all linear previsions on the set of gambles $\mathcal{L}(\mathcal{O})$. On the Bayesian sensitivity analysis interpretation, each linear prevision $P(\cdot|x)_i$ represents a so-called *random protocol* (Shafer, 1985), understood to be a protocol that selects an incomplete observation o from the set $\mathit{\Gamma}(x)$ of observations compatible with state x, with probability $p(o|x)$. The set $\mathcal{M}(\underline{P}(\cdot|x))$ contains all possible random protocols, and its lower envelop $\underline{P}(\cdot|x)$ models that there is no information to determine which random protocol is being followed.

MDI Irrelevance assumption We now can state de Cooman and Zaffalon (2004) *irrelevance assumption*, MDI. For all gambles f on $\mathcal{C}$

$$\underline{P}(f|x,o) = \underline{P}_0(f|x) \quad \text{for all } x \in \mathcal{X} \text{ and } o \in \mathit{\Gamma}(x).$$

The MDI assumption states that, conditional on the attributes variable X, the observations variable O is irrelevant to the class, or in other words, that the incomplete observations $o \in \mathit{\Gamma}(x)$ can influence our beliefs about the class only indirectly through the value x of the attributes variable X.

4.2 CAR/MAR vs MDI

Suppose we are studying the relationship between information sources, X, and hypotheses, H. We suppose that observing a value x for X in $\mathcal{X}$ affects which value we assign to H from $\mathcal{H}$. Suppose that we cannot directly observe or access the value for X. However, there is a protocol that allows us to make an incomplete observation O of X.

MDI tells us that if we have a precise observation that $X = x$, then the additional knowledge of the incomplete observation $O = o$ about x will not change our evaluation of H. Put another way, if you know the value of the precise observation, then knowing what incomplete observation this generates does not tell you anything new.

de Cooman and Zaffalon think that this characterizes the common, distinguishing feature of problems of missing data:

> ...when something that can be missing is actually measured, the problem of missing data disappears. Let us consider the opposite case, where the bare fact that an attribute is not measured is directly relevant to predicting the class. This fact should then become part of the classification

model by making a new attribute out of it, and treating it accordingly, so that this should not be regarded as a problem of missing information. Stated differently, once the model properly includes all the factors that are relevant to predicting the class, (MDI) follows naturally [1, p. 104].

Observe the following points:

CAR/MAR $\Rightarrow$ MDI CAR/MAR is stronger than MDI. So, when CAR is satisfied it follows that MDI is satisfied.

MDI $\not\Rightarrow$ CAR/MAR The converse obviously does not hold.

Attributes and Protocols MAR states that any incomplete observation o is equally likely to have been produced by all the attribute vectors x. Hence, MAR says something about the protocol or *incompleteness mechanism* that produces o. MDI, on the other hand, states that if one knows the attribute vector precisely, then knowing in addition what o produces won't give you any information. The important difference here, then, is that the classification of o as irrelevant depends on the attributes under MDI whereas it depends on the underlying *incompleteness mechanism* under MAR.

When MAR is justified There are cases where MAR is justified and MDI is too weak: example, the case of the always-missing-observation. MAR tells you that you can "marginalize out" this variable. However, de Cooman and Zaffalon are quite explicit in advising to use their updating rule *only* when nothing is known about the incompleteness mechanism. Yet, in this case, we do know something about the protocol. So, MAR is reasonable to hold.

Testing CAR/MAR and MDI CAR/MAR assumption cannot be tested statistically in the sense that we cannot use incomplete observations to check whether the assumption is reasonable. For the same reason MDI is not testable, either. However, this observation is a reason to favor MIDI to CAR/MAR:

> To understand this, let us, for the sake of simplicity, look at the case of precise probabilities: it should be tested whether or not $p(c|x, o) = p(c|x)$, for all classes c...The problem is that the precise observation x is always hidden to us; we can only see the incomplete observation o. So, in a statistical inference setting only $p(c, o)$ and not $p(c, o, x)$ would be accessible via the data. Therefore, there appears to exist a fundamental limitation of statistical inference in the presence of missing data: the actually observed data seem not to allow us to test our assumptions about the missing data mechanism, but nevertheless our inferences rely

heavily upon the specific assumptions that we make about it! This is one of the reasons why we are advocating that only those assumptions should be imposed that are weak enough to be tenable [1, p. 105].

5 Bayes Nets and Credal Nets

Bayesian nets themselves present a common case of updating from incomplete information, since typically the evidence set will not contain all attributes. It should be noted that MAR is the default choice for Bayes Nets. But it needn't be the only choice, however.

The main problem (that I see) facing a switch from Belief functions as the underlying mathematical framework to a theory based on lower previsions concerns how to effect inference within the more expressive language.

A Bayes net is defined by a DAG and a set of conditional probability mass functions. We observed that lower previsions is a more expressive language, so this representational scheme won't do. Instead, de Cooman and Zaffalon work with *credal nets*, which are DAGs composed of *credal sets*. A *credal set* (Levi 1980) is a DAG composed of sets of probability measures or, equivalently, linear lower prevision.

Viewed in terms of the set sets of probabilities, credal nets can have a huge number of extreme points; hence, calculating the lower and upper probabilities of a credal net is with strong extensions is NP-Hard *even for polytrees* (Ferreira da Rocha and Cozman, UAI Proceedings 2002. See note 22 in [1] for caveat). Bayes nets, by contrast, take polynomial time for polytrees.

deCooman and Zaffalon report that their updating rule for credal nets allows classification with credal nets to be realized with the same complexity needed for Bayesian nets. This is a significant result. Presumably, these rules will be quickly available to then reduce the search space of a credal net by listing the set of dominating possible optimal classes.

References

[1] Gert de Cooman and Marco Zaffalon. Updating beliefs with incomplete observations. *Artificial Intelligence*, 159:75–125, 2004.

[2] Bruno de Finetti. *Theory of Probability: A critical introductory treatment.* Wiley, 1974.

[3] Rolf Haenni and Stephan Hartmann. Modeling partially reliable information sources: a general approach based on Dempster-Shafer theory. Presented at FEW 2005, Austin, Texas, May 2005.

[4] Timothy Herron, Teddy Seidenfeld, and Larry Wasserman. Divisive conditioning: further results on dialation. *Philosophy of Science*, 64:411–444, 1997.

[5] Henry E. Kyburg, Jr. and Howard Smokler, editors. *Studies in Subjective Probability.* Wiley and Sons, New York, 1964.

[6] Isaac Levi. Consonance, dissonance and evidentiary mechanisms. In P. Gärdenfors, B. Hannsson, and N. Sahlin, editors, *Evidentiary value: Philosophical, Judicial and Psychological Ascpects of a Theory*, pages 27–43. Gleerups, 1983.

[7] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. Wiley, New York, 1987.

[8] Teddy Seidenfeld and Larry Wasserman. Dialation for sets of probabilities. *The Annals of Statistics*, 21:1139–154, 1993.

[9] Glenn Shafer. Conditional probability. *International Statistics Review*, 53:261–277, 1985.

[10] Glenn Shafer and Roger Logan. Implementing dempster's rule for hierarchical evidence. *Artificial Intelligence*, 33:271–298, 1987.

[11] Prakash Shenoy and Glenn Shafer. Propagating belief functions with local computations. In *IEEE Expert*, volume 1, pages 43–52, 1986.

[12] Philippe Smets and Robert Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.

[13] C. A. B. Smith. Consistency in statistical inference (with discussion). *Journal of the Royal Statistical Society*, 23:1–37, 1961.

[14] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[15] Peter Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996.