Measure-free conditioning, probability and non-monotonic reasoning

Didier Dubois   Henri Prade

Laboratoire Langages et Systemes Informatiques
Institut de Recherche en Informatique de Toulouse
Universite Paul Sabatier, 118 route de Narbonne
31062 TOULOUSE Cedex (FRANCE)

Abstract

Recent results in the foundations of probability theory indicate that a conditional probability can be viewed as a probability attached to a mathematical entity called a measure-free conditional. Such a measure-free conditional can receive a semantics in terms of a trivalent logic and logical operations are defined on conditionals in terms of truth-tables. It is shown that these results can be useful to justify Cox's axiomatic framework for probability as well as its application to other theories of uncertainty (Shafer's plausibility functions and Zadeh's possibility measures). Moreover it is shown that measure-free conditionals have the properties of well-behaved non-monotonic inference rules.

1 Introduction

Some analogy has been recently observed between Bayesian approaches to automated reasoning, and non-monotonic logics [Pearl, 1988, chap. 10]. Independently, the question of representing conditionals in accordance with conditional probability, that has puzzled philosophers in the seventies [Harper et al., 1981) is currently revived by mathematicians [Calabrese, 1987 ; Goodman and Nguyen, 1988] ; they have proposed an algebraic solution that contrasts with earlier attempts relying on possible world semantics and modal logics. At the same time, the debate between various numerical approaches to uncertainty in automated reasoning is still raging (see [Cheeseman, 1988] and the appended comments); the main defense of probabilistic orthodoxy seems to be founded on Cox [1946|'s axiomatic approach to conditional probability. This paper aims to put together these research trends and make the following observations : i) measure-free conditioning may be a good approach to the representation of inference rules (such as production rules in expert systems) which allow to capture non-monotonic features of commonsense reasoning ; basically, a conditional 'if b then a' denoted by alb, can be true, false or inapplicable in the proposed approach ; ii) Cox's system of conditional probability axioms is an example of homomorphism between a Boolean algebra augmented with measure-free conditionals and the unit interval ; however this homomorphism is not unique, and there is room for other non-additive measures of uncertainty, contrary to what is claimed by probability theory tenants ; iii) viewed as inference rules, measure-free conditionals satisfy the properties suggested by Gabbay [1985], as basic ones for non-monotonic reasoning systems ; these properties were already put forward by Adams [1975] ten years before in a probabilistic logic as recently advocated by Pearl [1988]. The three sections of this paper reflect these three points. What they suggest is that it may become possible to envisage a conjoint development of categorical and numerical approaches to reasoning under uncertainty, the latter being only a refinement of the former.

This work is partially supported by a grant from the PRC-GRECO Intelligence Artificielle.

2 A logical, Measure-Free View of Conditioning

The starting motivations of many models of conditionals which have been proposed do not refer to the notion of conditional probability and often these models are not fully compatible with probabilities. By contrast, Calabrese 11987], Goodman and Nguyen [1988] have tried to give a meaning to alb when the probability is removed. In the following, the symbol alb is informally interpreted as representing a production rule "if b then a", which means "when b is true then a can be added to the set of true facts otherwise the rule is not applicable".

Viewing alb as linking two propositions a and b of prepositional logic, semantics can be given to it under the form of an incomplete truth table, namely denoting t the truth-assignment function, t(alb) = 1 when a and b are both true, t(alb) = 0 when a is false and b is true. When b is false, alb is considered as inapplicable, which is denoted as t(alb) = ?. This symbol means that any truth value in {0,1] can be assigned to alb. Such semantics are in accordance with the usual meaning of production rules ; this proposal turns out to be exactly equivalent to a definition used by Schay [1968] twenty years ago and also explicitly appears in De Finetti's main paper on subjective probability [De Finetti, 1964]. The difference between b —> a and alb is easily expressed by means of their truth-tables : namely t(b —> a) = t(alb) when t(b) = 1 only.

t(b)	t(a)	t(b → a)	t(a | b)
1	1	1	1
1	0	0	0
0	1	1	?
0	0	1	?

Table 1 : Material implication versus conditioning symbol

This truth-functional definition separates interpretations (or possible worlds) into three classes : examples for the rule (t(alb) = 1), exceptions to the rule (t(alb) = 0) and irrelevant interpretations (t(alb) = ?). In [Dubois and Prade, 1985] we notice that t(alb) can be implicitly defined by means of the equation :

$$t(a \wedge b) = t(a|b) * t(b) \qquad (1)$$

where $*$ is the conjunction operation on $\{0,1\}$, defined by $1 * 1 = 1$ and $1 * 0 = 0 * 1 = 0 * 0 = 0$. Note that (1) obviously agrees with probabilities. Indeed if t(**b**) = 1 then t(**a**|**b**) coincides with Prob(**a**|**b**), and when t(**b**) = 0, Prob(**b**) = Prob($\mathbf{a} \wedge \mathbf{b}$) = 0.

Any proposition **x** such that $t(\mathbf{a} \wedge \mathbf{b}) = t(\mathbf{x}) * t(\mathbf{b})$ can stand for **a**|**b**. Note that filling the incomplete truth-table can only be done in four ways, equating **a**|**b** to one of the following propositions : $\mathbf{a} \wedge \mathbf{b}$, **a**, $\mathbf{a} \leftrightarrow \mathbf{b}$, $\mathbf{b} \rightarrow \mathbf{a}$ in a language containing only **a** and **b** as proposition symbols, and assuming that t(**a**) and t(**b**) can be independently valued ($\leftrightarrow$ is the equivalence symbol). **a**|**b** can be thus identified to a set of propositions and "|" does not define a connective in the usual sense.

At this point, it is interesting to recall the following definition, due to Goodman and Nguyen [1988], where **a**, **b**, **x** denote elements of a Boolean algebra $\mathcal{B}$

$$\forall\, \mathbf{a}, \mathbf{b} \in \mathcal{B},\ \mathbf{a}|\mathbf{b} = \{\mathbf{x}, \mathbf{x} \wedge \mathbf{b} = \mathbf{a} \wedge \mathbf{b}\} \qquad (2)$$

A slightly different definition which turns to be equivalent in the finite case has been independently proposed by Calabrese [1987]. It is easy to verify that (1) corresponds to the semantics of the above definition. Moreover $\mathbf{a}|\mathbf{b} = \{\mathbf{a} \wedge \mathbf{b} \le \mathbf{x} \le \neg\mathbf{b} \vee \mathbf{a}\}$ where $\le$ denotes the usual partial ordering in the Boolean algebra, expressing entailment, i.e. $\mathbf{b} \le \mathbf{a}$ if and only if $\neg\mathbf{b} \vee \mathbf{a} = \mathbb{1}$, where $\mathbb{1}$ denotes the greatest element in $\mathcal{B}$.

Considering a Boolean algebra $\mathcal{B}$ of propositions with tautology $\mathbb{1}$ and contradiction $\mathbb{O}$, we can define the set $\mathcal{B}|\mathcal{B} = \{\mathbf{a}|\mathbf{b}, (\mathbf{a}, \mathbf{b}) \in \mathcal{B}^2\}$. $\mathcal{B}$ can be identified as the subset $\{\mathbf{a}|\mathbb{1}, \mathbf{a} \in \mathcal{B}\}$. Indeed $t(\mathbf{a}|\mathbb{1}) = t(\mathbf{a})$ from (1). Note that in the language of expert systems a subset $\mathcal{K}$ of $\mathcal{B}|\mathcal{B}$ is a knowledge base, $\mathcal{K} \cap \mathcal{B}$ may be viewed as a factual base, and $\mathcal{K} \cap (\mathcal{B}|\mathcal{B} - \mathcal{B})$ as a rule base. Rules of the form $\mathbf{a}|\mathbb{O}$ are not very interesting since they are never applicable ($t(\mathbf{a}|\mathbb{O}) = ?$, $\forall \mathbf{a}$). Besides, $\mathbf{a}|\mathbf{a} = \partial(\mathbf{a}) = \{\mathbf{x}, \mathbf{a} \le \mathbf{x}\}$ and thus contains **a** and $\mathbb{1}$; thus **a**|**a** cannot be identified with $\mathbb{1}$. This is natural since the rule **a**|**a** is applicable only if **a** is known to be true, so that $t(\mathbf{a}|\mathbf{a}) \in \{1, ?\}$. More generally, if $\mathbf{a} \rightarrow \mathbf{b} = \mathbb{1}$ then $t(\mathbf{b}|\mathbf{a}) \neq 0$ only.

It can be checked using (2) that

$$\mathbf{a}|\mathbf{b} = \mathbf{c}|\mathbf{d} \text{ if and only if } \mathbf{a} \wedge \mathbf{b} = \mathbf{c} \wedge \mathbf{d} \text{ and } \mathbf{b} = \mathbf{d} \qquad (3)$$

as already noticed by De Finetti [1964]. The following equalities are worth pointing out

$$\mathbf{a}|\mathbf{b} = (\mathbf{a} \wedge \mathbf{b})|\mathbf{b} = (\mathbf{a} \leftrightarrow \mathbf{b})|\mathbf{b} = (\mathbf{b} \rightarrow \mathbf{a})|\mathbf{b} \qquad (4)$$

They are in agreement with our interpretation of **a**|**b** in terms of a production rule. Here we have four ways of describing the same rule.

Since $\mathbf{a}|\mathbf{b} = \{\mathbf{x}, \mathbf{a} \wedge \mathbf{b} \le \mathbf{x} \le \neg\mathbf{b} \vee \mathbf{a}\}$ and $\mathbf{c}|\mathbf{d} = \{\mathbf{y}, \mathbf{c} \wedge \mathbf{d} \le \mathbf{y} \le \neg\mathbf{d} \vee \mathbf{c}\}$ the following natural partial ordering can be defined on $\mathcal{B}|\mathcal{B}$ and will be also denoted by '≤'

$$\mathbf{a}|\mathbf{b} \le \mathbf{c}|\mathbf{d} \Leftrightarrow \mathbf{a} \wedge \mathbf{b} \le \mathbf{c} \wedge \mathbf{d} \text{ and } \neg\mathbf{b} \vee \mathbf{a} \le \neg\mathbf{d} \vee \mathbf{c} \qquad (5)$$

This definition was introduced by Goodman and Nguyen [1988]. Letting $\mathbf{b} = \mathbf{d} = \mathbb{1}$, it appears that this definition extends the usual entailment relation in $\mathcal{B}$. In the rule interpretation this ordering relation corresponds to an entailment relation. Indeed, it means that each time **a**|**b** is true **c**|**d** is true also, and that each time **c**|**d** is false, **a**|**b** is false too, that is any example of the rule **a**|**b** is also an example of the rule **c**|**d**, and any exception to **c**|**d** is an exception to **a**|**b**. It can be checked that (5) is compatible with (3), i.e. $\mathbf{a}|\mathbf{b} \le \mathbf{c}|\mathbf{d}$ and $\mathbf{a}|\mathbf{b} \ge \mathbf{c}|\mathbf{d} \Rightarrow \mathbf{a}|\mathbf{b} = \mathbf{c}|\mathbf{d}$.

For instance "generally, birds are small flying animals" entails that "generally, small birds fly", with **a** = "small and fly", **b** = "bird", **c** = "fly", **d** = "small bird". Rule **c**|**d** has more examples and less exceptions than **a**|**b**.

Besides, as pointed out in Dubois and Prade [1988], the relation ≤ defined by (5) corresponds in terms of truth-values, to the ordering $0 \le ? \le 1$, still using the same symbol "≤" between values of the truth function, i.e. $\mathbf{a}|\mathbf{b} \le \mathbf{c}|\mathbf{b}$ if and only if $t(\mathbf{a}|\mathbf{b}) \le t(\mathbf{c}|\mathbf{b})$ in all possible situations, and consequently $\mathbf{a}|\mathbf{b} = \mathbf{c}|\mathbf{d}$ if and only if their truth-tables coincide. In particular we have

$$\forall \mathbf{c},\ \mathbf{a}|\mathbf{b} \le (\mathbf{a} \vee \mathbf{c})|\mathbf{b} \qquad (6)$$

which means that if the rule a|b holds, any more imprecise conclusion a v c can be also produced. But,

there is no universal ordering between a|(b A C) *and* a|b (7)

More precisely, the rule a|b may be true or may be false while the rule $\mathbf{a}|(\mathbf{b} \wedge \mathbf{c})$ is not applicable. This contrasts with the material implication. Thus we capture a form of non-monotonicity in this framework.

The problem of extending operations such as negation, intersection and union to $\mathcal{B}|\mathcal{B}$ has been addressed by Schay [1968], Calabrese [1987], and Goodman and Nguyen [1988] sometimes in different ways. There is a consensus about negation, i.e.

$$\neg(\mathbf{a}|\mathbf{b}) = (\neg\mathbf{a}|\mathbf{b}) \qquad (8)$$

In terms of truth-values, it corresponds to extend the negation operation by postulating that $t(\neg\mathbf{a}) = ?$ if $t(\mathbf{a}) = ?$. That is to say $\neg(\mathbf{a}|\mathbf{b})$ corresponds to the converse rule "if b then not a". This is quite different from what happens with the material implication where $\neg(\mathbf{b} \rightarrow \mathbf{a}) = \mathbf{b} \wedge \neg\mathbf{a} \neq \mathbf{b} \rightarrow \neg\mathbf{a}$!

There exist three different proposals for defining the conjunction of (a|b) and (e|d), which may appear under various equivalent forms, since due to (4) there are at least four ways of describing the same rule.

In terms of truth-values it is pointed out in [Dubois and Prade, 1988) that these three definitions correspond to three possible extensions of the binary conjunction operation which preserve the symmetry and which take into account the symbol ? introduced in Table 1. Namely, the three different conjunctions, denoted by $\wedge$, $.$, $\cap$ are defined by

$$\forall\, x \in \{0, ?, 1\},\ x \wedge ? = x = ? \wedge x\ ;\ x\,.\,? = \min(x,?) = ?\,.\,x\ ;\ x \cap ? = ? = ? \cap x \qquad (9)$$

The first conjunction is such that the combination of something true (resp. false) with something inapplicable is true (resp. false). The second conjunction is defined in agreement with the ordering $0 \le ? \le 1$. Note that these conjunction operations were first considered in the framework of trivalent logics, by Sobocinski, Lukasiewicz and Bochvar respectively ; see Rescher [1969]. Using the same notation for combining piopositions or their truth values, the first conjunction is equivalent to the following definition considered by Schay [1968| and Calabrese [1987] (under other equivalent forms due to (4))

$$(\mathbf{a}|\mathbf{b}) \wedge (\mathbf{c}|\mathbf{d}) = [(\mathbf{b} \rightarrow \mathbf{a}) \wedge (\mathbf{d} \rightarrow \mathbf{c})] \,|\, (\mathbf{b} \vee \mathbf{d}) \qquad (10)$$

From the point of view of rule-based systems, ^ means that the two rules a|b and e|d are available and form a rule base. It is natural to define the applicability of a rule base $\{\mathbf{a}_i|\mathbf{b}_i, i = 1,n\}$ to a factual base $\mathcal{F}$ whenever at least one rule i is applicable to

$\mathcal{F}$, i.e. $\vee_{i=1,n} b_i$ is true. The two other conjunctions correspond to different points of view ; see [Dubois and Prade, 19881. Their expressions are

$$(a|b) \,.\, (c|d) = (a \wedge c) \mid [(\neg a \wedge b) \vee (\neg c \wedge d) \vee (b \wedge d)] \quad (11)$$

$$(a|b) \cap (c|d) = (a \wedge c) \mid (b \wedge d) \quad (12)$$

and they have been considered by Goodman and Nguyen [1988] and by Schay [1968] respectively. Note that

$$(a|b) \curlywedge (c|b) = (a|b) \,.\, (c|b) = (a|b) \cap (c|b) = (a \wedge c)|b$$

so as the three conjunctions differ only when b # d. (11) and (12) are not very natural when alb is viewed as a rule. For instance (12) assumes that a rule base containing alb and eld is applicable when <u>both</u> rules are applicable.(10), (11) and (12) basically differ by the conditions in which they are applicable, and the advantage of (10) is that it is the least demanding in that respect.

Three disjunction operations can be derived by duality using De Morgan laws. Lastly, the conditioning process can be iterated in such a way that $\mathfrak{B}|\mathfrak{B}$ remains closed under this process. Calabrese [1987] has proposed the following definition : $(a|b)|c = a|(b \wedge c)$; it corresponds to the convention ?I0 = ? = ?II (using the same symbol T for propositions and for values of the truth function). Two meanings can be envisaged for the companion expression al(blc) namely i) al(blc) = al(b A c) stating 11? = ? = 01?, or ii) a l(blc) = al(c -> b) (this is Calabrese's definition) and corresponds to set 1l? = 1 and 01? = 0. See [Dubois and Prade, 1988] for other results and discussions.

Schay [1968] indicates that $\curlywedge$ and its dual $\curlyvee$ are not distributive with respect to each other so that $(\mathfrak{B}|\mathfrak{B}, \curlywedge, \curlyvee, \neg)$ is no longer a Boolean algebra. See also Calabrese [1987].

3 About Cox's Axioms

As recalled by Cheeseman [1988], Cox [1946] has postulated the following axioms that degrees of "reasonable expectation" valued on the unit interval should satisfy :

A1 : $P(\neg a|b) = S(P(a|b))$

A2 : $P(a \wedge b|c) = F(P(a|c), P(b|a \wedge c))$

A3 : S is stricdy decreasing ; F is strictly increasing in each place ; S and F are continuous.

Using these assumptions only, it is possible to prove that S(x) = 1 - x up to an isomorphism, that F is associative (due to the Boolean structure of the set of propositions a, b...), and is a product up to an isomorphism ; so that the set function P is a probability measure. See also Aleliunas [1988] for a slightly different approach to the same question.

Shafer [1988] has criticized A2 as being natural only for someone who is familiar with the usual definition of conditional probability. For instance conditional probability can also be defined as

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} = \frac{1}{1 + \dfrac{P(\neg a \wedge b)}{P(a \wedge b)}}$$

and it may look natural to start with the "natural" requirement that P(alb) be defined as a function of $P(a \wedge b)$ and $P(\neg a \wedge b)$ only, since it reflects the relative strength of a and -a in the environment where b is true. It would lead to an axiom

A'2 : $P(a|b) = F(P(a \wedge b), P(\neg a \wedge b))$

which explicitly defines P(alb) while A2 only provides an implicit definition. It is hard to choose between A2 and A'2 as to which is the most natural axiom !

It turns out that A1 and A2 take full meaning not only as axioms postulated from pure commonsense, but as expressing an homomorphism between the structure $\mathfrak{B}|\mathfrak{B}$ described in section 1 and the unit interval. Indeed, measure-free conditionals satisfy the following property, which is easy to prove from (10) and (4).

$$(a \wedge b)|c = [a|c] \curlywedge [b|(a \wedge c)] \quad (13)$$

A similar identity holds with the conjunction defined by (11) but not with the one defined by (12). So that A2 becomes natural as a compositionality assumption with respect to the extended conjunction of conditionals of the form ale and bl(a A C).

Axiom Al clearly requires the negation of a measure-free conditional as it appears in (8). However this axiom also presupposes that the meaning of the extreme values of P(alb) (i.e. 0 and 1) is well-understood. This is clearly a matter of convention. Cox's convention is that 1 means certainty (Probability = 1) and 0 means impossibility (Probability = 0). Axiom Al becomes very natural since it means (along with A3) the more probable a, the less probable - a . However, another convention is reasonable as well, namely 1 means possibility (i.e. consistency with available knowledge) and 0 means impossibility. Under this new convention, A1 does not sound reasonable at all, since it would mean, along with A3 : the more possible a, the less possible —a. But in case of incomplete Knowledge, one may find that a and -a are equally and totally possible. A more natural substitute to Al would be : the more impossible —>a, the more certain a. In other words, when P(alb) ranges from impossibility (0) to possibility (1), S(P(alb)) = P(-alb) does not qualify -a in the same way : P(->alb) = 1 means that -a is certain while P(-alb) = 0 means that —a is totally <u>uncertain</u> (i.e. it corresponds to a state of ignorance). Hence changing the meaning of the end-points of the unit interval may lead to drop axiom A l, and to consider two set-functions, one for possibility, say II, one for certainty say C, that exchange via the duality property

$$\Pi(a|b) = S(C(\neg a|b)) \quad (14)$$

and it that may act as a substitute to A1.

Axiom A3 is technical, and was stated in a stronger form by Cox [1946], originally. It is remarkable that only probability measures emerge as the unique solution to A1-A3. However one must be aware that if A3 is further relaxed by requiring that F be strictly isotonic only, i.e.

$$x > x', y > y' \Rightarrow F(x,y) > F(x',y')$$

then, A1-A3 have solutions which are not probability measures. Indeed F = minimum is isotonic, and there exist set-functions g such that g(alb) = 1 - g(-alb), g(a A b) = min(g(alb), g(b)), $\forall a, b$ such as the following one on a 4-element set Ω = {1,2,3,4}

g({1}) = 0.7	g({1,2}) = .75	g({2,3}) = .3	g({2,3,4}) = .3
g({2}) = 0.3	g({3,4}) = .25	g({1,4}) = .7	g({1,3,4}) = .7
g({3}) = 0.2	g({1,3}) = .7	g(∅) = 0	g({1,2,4}) = .8
g({4}) = 0.1	g({2,4}) = .3	g(Ω) = 1	g({1,2,3}) = .9

They are such that $\forall \mathbf{a}, \forall \mathbf{b}$, $g(\mathbf{a} \wedge \mathbf{b}) < g(\mathbf{b})$ and $g(\neg\mathbf{a} \wedge \mathbf{b}) < g(\mathbf{b}) \Rightarrow g(\mathbf{a} \wedge \mathbf{b}) = g(\neg\mathbf{b} \vee \mathbf{a}) = g(\mathbf{a})$). See Dubois and Prade [1988] for more discussions on this point, g is monotonia under inclusion, but not decomposable through disjunctions of mutually exclusive propositions.

Another important issue is the compatibility between conditional probability and the entailment relation < between rules a|b. The following result can be established [Dubois and Prade, 1988) :

Proposition 1 : a|b < e|d implies P(a|b) < P(c|d), when P is a probability measure.

Proof : Let g be a monotonic function, i.e. such that $\mathbf{a} \le \mathbf{b}$ implies $g(\mathbf{a}) \le g(\mathbf{b})$. Let $\mathbf{a|b} \le \mathbf{c|d}$; then $g(\mathbf{a} \wedge \mathbf{b}) \le g(\mathbf{c} \wedge \mathbf{d})$ and $g(\neg\mathbf{b} \vee \mathbf{a}) \le g(\neg\mathbf{d} \vee \mathbf{c})$ due to the monotonicity of g ; the latter inequality also writes $g(\mathbf{b} \wedge \neg\mathbf{a}) \ge g(\mathbf{d} \wedge \neg\mathbf{c})$, since $\mathbf{d} \wedge \neg\mathbf{c} \le \mathbf{b} \wedge \neg\mathbf{a}$. Now $P(\mathbf{a|b}) = f(\frac{P(\neg\mathbf{a} \wedge \mathbf{b})}{P(\mathbf{a} \wedge \mathbf{b})})$ with $f(x) = 1/(1 + x)$. The above inequalities imply $\frac{P(\neg\mathbf{a} \wedge \mathbf{b})}{P(\mathbf{a} \wedge \mathbf{b})} \ge \frac{P(\neg\mathbf{c} \wedge \mathbf{d})}{P(\mathbf{c} \wedge \mathbf{d})}$, hence $P(\mathbf{a|b}) \le P(\mathbf{c|d})$. Q.E.D.

This result can be extended to Shafer [1976]'s plausibility measures, as well as possibility measures [Zadeh, 1978), [Dubois and Prade, 1988b]. Indeed let us assume that II is a plausibility measure, and that II satisfies A2 and A3. Note that F = product is once again the unique operation for the definition of conditioning, up to an isomorphism (e.g. Aczel [1966]). A2 corresponds to Shafer [19761's definition of conditional plausibility, in accordance with Demspter rule of conditioning. Now the extension of proposition 1 writes :

Proposition 2 : $\mathbf{a|b} \le \mathbf{c|d}$ implies $\Pi(\mathbf{a|b}) \le \Pi(\mathbf{c|d})$ where Π is a plausibility measure or a possibility measure.

Proof : it is enough to express

$$\Pi(\mathbf{a|b}) = \frac{\Pi(\mathbf{a} \wedge \mathbf{b})}{\Pi(\mathbf{b})} \text{ as } \frac{\Pi(\mathbf{a} \wedge \mathbf{b})}{\Pi(\mathbf{a} \wedge \mathbf{b}) + K(\mathbf{a},\mathbf{b})}$$

where $K(\mathbf{a},\mathbf{b}) = \sum\{m(\mathbf{e}) | \mathbf{e} \wedge \mathbf{b} \ne \mathbb{O}, \mathbf{e} \wedge \mathbf{a} \wedge \mathbf{b} = \mathbb{O}\}$, and m denotes the basic probability assignment defining Π. $K(\mathbf{a},\mathbf{b})$ is the weight bearing on exceptions to the rule $\mathbf{a|b}$. If $\mathbf{a|b} \le \mathbf{c|d}$ then $\Pi(\mathbf{a} \wedge \mathbf{b}) \le \Pi(\mathbf{c} \wedge \mathbf{d})$, and $K(\mathbf{a},\mathbf{b}) \ge K(\mathbf{c},\mathbf{d})$ (see [Dubois and Prade, 1988] for details). When Π is a possibility measure, the result also holds since Π is a plausibility measure such that $\Pi(\mathbf{a} \vee \mathbf{b}) = \max(\Pi(\mathbf{a}), \Pi(\mathbf{b}))$, $\forall \mathbf{a},\mathbf{b}$. A direct proof is to use the one of proposition 1 changing $f(x) = 1/(1 + x)$ into $f(x) = \min(1, 1/x)$. Q.E.D.

As a consequence, plausibility and possibility measures, although violating axiom AI (due to a matter of convention) define homomorphisms between $\mathfrak{B}|\mathfrak{B}$ and the unit interval. Since $\mathbf{a|b} \le \mathbf{c|d} \Leftrightarrow \neg\mathbf{a|b} \ge \neg\mathbf{c|d}$, the dual measure of certainty (belief measure or necessity measure) also satisfies Proposition 2, namely $\mathbf{a|b} \le \mathbf{c|d} \Rightarrow C(\mathbf{a|b}) \le C(\mathbf{c|d})$. Proposition 2 is an algebraic justification of Dempster's rule of conditioning. However several problems are still pending such as the extension of proposition 2 to more general upper and lower probability functions, and more generally the independence between axioms A2, A3 and the compatibility property expressed by proposition 2, when AI is dropped.

4 Measure-Free Conditionals as Default Rules

Several authors, and especially Pearl [1988] have suggested that a default rule such as "generally, b's are a's" could be interpreted as Probability(a|b) = HIGH. It may be tempting to consider the measure-free conditional a|b as a model of default, in the spirit of Reiter [1980]'s logic, i.e. without refering to any statistical interpretation, considering that the statistical component is carried by the probability attached to the conditional, and not by the conditional itself.

There are two conditions under which a|b represents a default:

-) there exists at least one interpretation for which t(a|b) = 1 (otherwise $\mathbf{a} \wedge \mathbf{b} = \mathbb{O}$ and it makes no sense to assert a|b)

-) there may exist exceptions to the rule, i.e. interpretations for which t(a|b) = 0 (otherwise a < b and a|b is nothing but a standard monotonic inference rule ; particularly, for any probability measure P, P(a|b) = 1, while here we wish to allow for $P(\mathbf{a|b}) \in (0,1)$).

Asserting "a|b" can thus be interpreted as : $\mathbf{b} \ne \mathbb{O}$ and $(\mathbf{a|b}) \ne (\mathbb{O}|\mathbf{b})$; these two conditions hold if and only if $\mathbf{a} \wedge \mathbf{b} \ne \mathbb{O}$, so that a|b means "there are examples of b's that are a's", i.e. the weakest kind of default rule one may apparently think of.

As a next step, it seems possible to use the ordering relation < between defaults and the conjunction ^ for the definition of a consequence relation. A default e|f can be deduced from {a|b,c|d} if and only if

$$(\mathbf{a|b}) \wedge (\mathbf{c|d}) \le (\mathbf{e|f}) \qquad (15)$$

(15) means the following : any example of one of the rules a|b or e|d, that is not an exception to the other rule is an example of e|f; and any exception to e|f is an exception to one of the rules a|b, or e|d. Particularly $(\mathbf{a|b}) \wedge (\mathbf{c|d}) \le (\mathbf{a|b})$ does not hold because an example of e|d can be simply irrelevant for a|b (i.e. $t(\mathbf{c} \wedge \mathbf{d}) = 1$ and t(b) = 0 so that $t((\mathbf{a|b}) \wedge (\mathbf{c|d})) = 1$ while t(a|b) = ?. On the contrary a rule e|f that satisfies (15) takes into account both rules since its examples are at least all those of each rule when the two rules do not contradict each other. This remark suggests that (15) defines e|f as a weak substitute to the set of defaults {a|b, e|d) in which both rules are still acting. By contrast we have $(\mathbf{a|b}) \cdot (\mathbf{c|d}) \le (\mathbf{a|b})$ with the conjunction defined by (11). Hence using this inequality as an inference rule does not look proper.

The following properties of the consequence relation < are noticeable, and easily checked using truth-tables :

$$(\mathbf{a|b}) \wedge (\mathbf{c|b}) \le \mathbf{c}|(\mathbf{a} \wedge \mathbf{b}) \qquad (16)$$
$$(\mathbf{a|b}) \wedge [\mathbf{c}|(\mathbf{a} \wedge \mathbf{b})] \le \mathbf{c|b} \qquad (17)$$
$$(\mathbf{c|a}) \wedge (\mathbf{c|b}) = \mathbf{c}|(\mathbf{a} \vee \mathbf{b}) \qquad (18)$$

(16) holds with the two other conjunctions as well since $(\mathbf{a} \wedge \mathbf{c})|\mathbf{b} \le \mathbf{c}|(\mathbf{a} \wedge \mathbf{b})$; (17) which is a direct consequence of (13) and (6) holds also for the conjunction defined by (11) ; by contrast (18) does not hold with any of the two other conjunctions. These relations can serve as inference rules that produce new defaults from existing ones. In Adams [1975]'s conditional logic a|b is interpreted as $P(\mathbf{a|b}) \ge 1 - \varepsilon$ where e is arbitrarily close to 0 and denoted b a. This interpretation is much more demanding than ours. However Adams [1975] found inference rules that are exactly (16-18), namely :

triangularity	$\mathbf{b} \to \mathbf{a}$	$\mathbf{b} \to \mathbf{c}$	$\Rightarrow$	$(\mathbf{b} \wedge \mathbf{a}) \to \mathbf{c}$
Bayes	$\mathbf{b} \to \mathbf{a}$	$(\mathbf{a} \wedge \mathbf{b}) \to \mathbf{c}$	$\Rightarrow$	$\mathbf{b} \to \mathbf{c}$

disjunction $\mathbf{a} \rightarrow \mathbf{c} \quad \mathbf{b} \rightarrow \mathbf{c} \quad \Rightarrow \quad (\mathbf{a} \vee \mathbf{b}) \rightarrow \mathbf{c}$

The rules are used by Pearl [1988] to build a probabilistic-like default logic. Starting from purely logical assignments, Gabbay [1985] proposed several axioms a non-monotonic deduction operation should satisfy, and especially :

$$\frac{\mathbf{b} \mid\!\sim \mathbf{a} \ ; \ \mathbf{b} \mid\!\sim \mathbf{c}}{\mathbf{b}, \mathbf{a} \mid\!\sim \mathbf{c}} \text{ (restricted monotonicity)} \quad \text{and} \quad \frac{\mathbf{b} \mid\!\sim \mathbf{a} \ ; \ \mathbf{a}, \mathbf{b} \mid\!\sim \mathbf{c}}{\mathbf{b} \mid\!\sim \mathbf{c}} \text{ (transitivity)}$$

Clearly, restricted monotonicity becomes triangularity in the probabilistic setting and correspond to (16), while (17), related to Bayes rule, is simply a transitivity property that deduction must satisfy. These remarks suggest that a non-monotonic logic where defaults are modelled by measure-free conditionals is likely to have all the properties that a well-behaved non-monotonic logic should satisfy, especially the possibility to infer new defaults, and the reasoning by cases (due to (18)). In probabilistic terms, alb means P(alb) > e where £ is positive but can be arbitrarily small. Thus it is more general than Adams [1975]'s interpretation ; (16-18) then translate into the following inequalities:

$$P(\mathbf{c}|\mathbf{a} \wedge \mathbf{b}) \geq P(\mathbf{a} \wedge \mathbf{c}|\mathbf{b}) \quad ; \quad P(\mathbf{c}|\mathbf{b}) \geq P(\mathbf{a}|\mathbf{b}) \,.\, P(\mathbf{c}|\mathbf{a} \wedge \mathbf{b})$$

$$P(\mathbf{c}|\mathbf{a}) > \varepsilon \,,\, P(\mathbf{c}|\mathbf{b}) > \varepsilon' \quad \Rightarrow \quad P(\mathbf{c}|\mathbf{a} \vee \mathbf{b}) > \varepsilon''.$$

5 Conclusion

This paper is meant to investigate some consequences of recent results about measure-free conditioning. Our contention is that measure-free conditionals could be a good way of modeling non-monotonic production rules in accordance with numerical theories of uncertainty ; in other words this approach equips uncertain rules in expert systems with clear semantics. However the use of uncertainty coefficients is not compulsory : propositional logic augmented with measure-free conditionals can be used as a formal system of non-monotonic logic. Properties of this system will be investigated in the future.

References

[Aczel, 1966] Aczel J. Lectures on Functional Equations and their Applications. Academic Press, New York, 1966.

[Adams, 1975] Adams E.W. The Logic of Conditionals. D. Reidel, Dordrecht, 1975.

[Aleliunas, 1988] Aleliunas R. A new normative theory of probabilistic logic. In Proceedings of the 7th Biennial Conference of the Canadian Society for Computational Studies of Intelligence (R. Goebel, ed.), pages, 67-74, Edmonton, Alberta, Canada, June 6-10, 1988.

[Calabrese, 1987) Calabrese P. An algebraic synthesis of the foundations of logic and probability. Information Sciences. 42 : 187-237, 1987.

[Cheeseman, 1988] Cheeseman P. An inquiry into computer understanding (with discussions). Computational Intelligence. 4(1): 57-142, 1988.

|Cox, 1946] Cox R.T. Probability, frequency and reasonable expectation. Amer. J. Phys.. 14 : 1-13, 1946.

[De Finetti, 1964] De Finetti B. La prevision, ses lois logiques et ses sources subjectives. Ann. Inst. H. Poincare 7. 1937, translated by H. Kyburg Jr., in Studies in Subjective Probability (H. Kyburg Jr., H. Smokier, eds.), pages 95-158, Wiley, New York, 1964.

[Dubois and Prade, 1985] Dubois D. Prade H. Theorie des Possibilites. Applications a la Representation des Connaissances en Informatique. Masson, Paris, 1985. 2nd edition, pages 135-138 and pages 140-143, 1987.

[Dubois and Prade, 1988] Dubois D. Prade H. The logical view of conditioning and its applications to possibility and evidence theories. In Report L.S.I. n° 301, Universite Paul Sabatier, Toulouse, April 1988. To appear in Int. J. of Approximate Reasoning. 4, 1990.

[Dubois and Prade, 1988b] Dubois D. Prade H. Possibility Theory - An Approach to Computerized Processing of Uncertainty. Plenum Press, New York, 1988.

[Gabbay, 19851 Gabbay D.M. Theoretical foundations for non-monotonic reasoning in expert systems. In Logics and Models of Concurrent Systems (K.R. Apt, ed.), pages 439-457, Springer Verlag, Berlin, 1985.

[Goodman and Nguyen, 1988] Goodman I.R. Nguyen H.T. Conditional objects and the modeling of uncertainties. In Fuzzy Computing (M.M. Gupta, T. Yamakawa, eds.), pages 119-138, North-Holland, Amsterdam, 1988.

[Harper et al., 1981) Harper W.L. Stalnaker R. Pearce G., editors. Ifs - Conditionals. Belief. Decision. Chance and Time. D. Reidel, Dordrecht, 1981.

[Pearl, 1988] Pearl J. Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference. Morgan Kaufmann, San Mateo, 1988.

[Reiter, 1980] Reiter R. A logic for default reasoning. Artificial Intelligence. 13 : 81-132, 1980.

[Rescher, 1969| Rescher N. Many-Valued Logic. McGraw-Hill, New York, 1969.

[Schay, 1968] Schay G. An algebra of conditional events. Journal of Mathematical Analysis and Applications. 24 : 334-344, 1968.

[Shafer, 1988] Shafer G. Comments on An inquiry into computer understanding by Peter Cheeseman. Computational Intelligence (Canada). 4 : 121-124, 1988.

[Shafer, 1976] Shafer G. A Mathematical Theory of Evidence. Princeton Univ. Press, New Jersey, 1976.

[Zadeh, 1978] Zadeh L.A. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1 : 3-28, 1978.