COHERENCE OF RULES FOR DEFINING CONDITIONAL POSSIBILITY

PETER WALLEY AND GERT DE COOMAN

ABSTRACT. Possibility measures and conditional possibility measures are given a behavioural interpretation as marginal betting rates against events. Under this interpretation, possibility measures should satisfy two consistency criteria, known as 'avoiding sure loss' and 'coherence'. We survey the rules that have been proposed for defining conditional possibilities and investigate which of them satisfy our consistency criteria in two situations of practical interest. Only two of these rules satisfy the criteria in both cases studied, and the conditional possibilities produced by these rules are highly uninformative. We introduce a new rule that is more informative and is also coherent in both cases.

1. INTRODUCTION

When the term 'possibility measure' was coined by Zadeh [29] for what is essentially a supremum preserving set function, the significance of this type of set function had already been recognised in a number of other contexts [21, 22, 23]. These set functions are especially simple to work with because they can be completely characterised by point functions. In this respect, they are similar to probability measures.

Consider a possibility space Ω, which is a non-empty set of possible states of the world. In this paper we assume that Ω is *finite*. A *possibility measure* $\overline{P}$ on Ω is a function, defined on the power set of Ω and taking values in the real unit interval $[0,1]$, such that $\overline{P}(A \cup B) = \max\{\overline{P}(A), \overline{P}(B)\}$, for all $A \subseteq \Omega$ and $B \subseteq \Omega$. Define the function $\pi\colon \Omega \to [0,1]$, such that $\pi(\omega) = \overline{P}(\{\omega\})$, $\omega \in \Omega$. The point function π then completely determines the set function $\overline{P}$, since for any $A \subseteq \Omega$, $\overline{P}(A) = \max\{\pi(\omega)\colon \omega \in A\}$. It is called the *possibility distribution* of $\overline{P}$, and it plays a similar role to a probability mass or density function.

Possibility measures have received much attention in fuzzy set theory [4, 5, 6, 12, 13, 29], mainly because of Zadeh's *possibility assignment equation* [29], a link which is claimed to exist between fuzzy sets and possibility distributions. Using this link, possibility measures are taken to model the uncertainty conveyed by vague and non-specific statements. For recent evaluations of this connection, we refer to [3, 25, 27].

But the relevance of possibility measures is certainly not restricted to fuzzy set theory alone. They constitute an important special class of *upper probabilities* in the behavioural theory of imprecise probabilities [24], as we argued in [8, 25, 26]. Provided a possibility measure $\overline{P}$ is *normal*, i.e. $\overline{P}(\Omega) = \max\{\pi(\omega)\colon \omega \in \Omega\} = 1$, its values $\overline{P}(A)$ can be given a behavioural interpretation as marginal rates for betting against the events A, where $A \subseteq \Omega$; see Section 2 for more details.

Conditioning is an important aspect of any uncertainty model. In the case of possibility measures, various rules have been proposed for defining conditional possibility measures or distributions from unconditional ones. The existence of so many different proposals for defining conditional possibilities, and the absence of a convincing justification for any of the rules or even a criterion for choosing between them, is disturbing. To evaluate the rules, we need (a) a specific interpretation of degrees of possibility, and (b) some consistency criteria to relate conditional and unconditional possibilities. In this paper we adopt the behavioural interpretation of possibility measures. This interpretation has the advantage that it leads directly to consistency criteria of avoiding sure loss and coherence, and hence it provides a way of

---

Date: 3 December 1998.

Key words and phrases. Coherence, possibility measure, conditional possibility, upper probability, imprecise probabilities, Dempster's rule, natural extension, possibility theory.

This paper is accepted for publication in the *International Journal of Approximate Reasoning*, 1998.

discriminating between the proposed conditioning rules according to whether they satisfy the criteria[1]. We also propose a new conditioning rule which appears to be superior to the others in several ways.

We shall be concerned with conditioning on variables rather than events, but the ideas of the paper apply equally well to the latter type of conditioning. To fix the notation, consider two variables X and Y. Let $\mathcal{X}$ and $\mathcal{Y}$ denote their respective sets of possible values. We assume that $\mathcal{X}$ and $\mathcal{Y}$ are both *finite*: the coherence conditions are more complicated, and more controversial, for infinite spaces. Write $\pi(\cdot,\cdot)$ for the unconditional possibility distribution, $\pi(x)$ and $\pi(y)$ for the marginal possibilities[2], and $\pi(\cdot|x)$ and $\pi(\cdot|y)$ for the conditional possibility distributions. Let

$$\begin{aligned}\overline{P}(A) &= \max\{\pi(x,y)\colon (x,y)\in A\} \quad \text{for non-empty } A\subseteq \mathcal{X}\times\mathcal{Y}\\ \overline{P}(B|y) &= \max\{\pi(x|y)\colon x\in B\} \quad \text{for non-empty } B\subseteq \mathcal{X},\ y\in\mathcal{Y}\\ \overline{P}(C|x) &= \max\{\pi(y|x)\colon y\in C\} \quad \text{for non-empty } C\subseteq \mathcal{Y},\ x\in\mathcal{X}\end{aligned}$$

with $\overline{P}(\emptyset)=\overline{P}(\emptyset|y)=\overline{P}(\emptyset|x)=0$, denote the *upper probability measures* or *possibility measures* generated by the possibility distributions $\pi(\cdot,\cdot)$, $\pi(\cdot|y)$ and $\pi(\cdot|x)$. The marginal distributions are given by

$$\begin{aligned}\overline{P}(\mathcal{X}\times\{y\}) &= \max\{\pi(u,y)\colon u\in\mathcal{X}\} = \pi(y)\\ \overline{P}(\{x\}\times\mathcal{Y}) &= \max\{\pi(x,v)\colon v\in\mathcal{Y}\} = \pi(x).\end{aligned}$$

The paper is organised as follows. In Section 2 we explain and discuss the behavioural interpretation of both unconditional and conditional possibility measures. In Sections 3 and 4, we introduce and define the consistency criteria of normality, avoiding sure loss and coherence, and we argue that it is indeed reasonable to impose them if we want our possibilistic models to reflect rational behaviour. We find that normality and avoiding sure loss are very easily satisfied in the context of possibility measures. The criterion of coherence is given more attention in Sections 5 and 6, where it is also used to discriminate between the various conditioning rules extant in the literature.

The definitions of these rules are listed below, together with a short discussion of their backgrounds. A more detailed overview and discussion can be found in [5]. It will be assumed here that the joint distribution $\pi(\cdot,\cdot)$ is normal, i.e. that $\max\{\pi(x,y)\colon (x,y)\in\mathcal{X}\times\mathcal{Y}\}=1$.

1.1. **Zadeh's rule.** Zadeh was the first to consider conditioning for possibility measures [29]. His rule is very simple, and consists in equating conditional degrees of possibility with unconditional ones:

$$\pi_{ZA}(x|y) = \pi(x,y), \quad x\in\mathcal{X},\ y\in\mathcal{Y}.$$

It has the disadvantage that it may produce unnormalised conditional possibility distributions $\pi_{ZA}(\cdot|y)$, even when the joint distribution is normal. Indeed, for any $y\in\mathcal{Y}$, $\pi_{ZA}(\cdot|y)$ is unnormalised whenever the marginal $\pi(y)<1$!

1.2. **Hisdal's equation.** Inspired by Bayes' rule, which can be written as $P(x,y)=P(x|y)P(y)$, Hisdal [16] proposed to define conditional possibilities using the following equation relating the conditional, marginal and joint distributions:

$$\pi(x,y) = \min\{\pi(x|y),\pi(y)\}, \quad x\in\mathcal{X},\ y\in\mathcal{Y}. \tag{1}$$

Whereas Bayes' rule leads to a unique value $P(x,y)/P(y)$ for the conditional probability $P(x|y)$ whenever $P(y)>0$, Hisdal's equation is much less restrictive. It has a unique solution $\pi(\cdot|y)$ only if $\pi(y)=1$. All its solutions are given by:

$$\pi(x|y) \in \begin{cases}\{\pi(x,y)\} & \text{if } \pi(x,y)<\pi(y)\\ [\pi(x,y),1] & \text{if } \pi(x,y)=\pi(y).\end{cases} \tag{2}$$

[1] The method used in this paper could be applied more generally to evaluate other kinds of rules for combining or modifying possibility distributions, by investigating whether the possibility distributions produced by the rules are coherent with the initial distributions. See pp. 45–46 of [25] for more details.

[2] We do not distinguish in notation between the marginal possibility distributions for X and Y. It will at all times be clear from the context which marginal is intended.

An arbitrary solution $\pi(\cdot|y)$ of (1) need not be normal, as is exemplified by the fact that Zadeh's rule picks out the *smallest* solution. There are a number of other conditioning rules which yield particular *normal* solutions to Hisdal's equation. We will consider the rules proposed by Ramer and by Dubois and Prade.

1.3. **Ramer's rule.** For a given y in $\mathcal{Y}$, Hisdal's equation may have more than one solution $\pi(x|y)$ only for those $x \in \mathcal{X}$ which maximise $\pi(\cdot, y)$, i.e. which satisfy $\pi(x, y) = \pi(y)$. Ramer's rule [18] consists in picking *one* x_o such that $\pi(x_o, y) = \pi(y)$, letting $\pi(x_o|y) = 1$, and $\pi(x|y) = \pi(x, y)$ for all other x in $\mathcal{X}$. It produces normal $\pi(\cdot|y)$, but it has the disadvantage of requiring an arbitrary choice whenever there is more than one x that maximises $\pi(\cdot, y)$.

1.4. **The Dubois-Prade rule.** The conditioning rule proposed by Dubois and Prade [10, 12, 13, 14] is a variation on this theme, but it does not suffer from the ambiguity in Ramer's rule, because it takes $\pi(x|y) = 1$ for *every* x which maximises $\pi(\cdot, y)$, i.e. $\pi(x, y) = \pi(y)$. It is defined as follows, for $x \in \mathcal{X}$, $y \in \mathcal{Y}$:

$$\pi_{DP}(x|y) = \begin{cases} \pi(x, y) & \text{if } \pi(x, y) < \pi(y) \\ 1 & \text{if } \pi(x, y) = \pi(y). \end{cases}$$

For every $y \in \mathcal{Y}$, this rule yields the *greatest*, or as Dubois and Prade call it, the *least specific* solution $\pi(\cdot|y)$ of Hisdal's equation, and $\pi(\cdot|y)$ is always normal.

1.5. **Dempster's rule.** In an important paper about upper and lower probabilities induced by multi-valued mappings [9], Dempster proposed a rule for conditioning upper probabilities. Since possibility measures are upper probabilities induced by a particular type of multi-valued mapping [7, 15, 22], Dempster's rule can be applied to the problem of defining conditional possibilities. This yields, for $x \in \mathcal{X}$, $y \in \mathcal{Y}$:

$$\pi_{DE}(x|y) = \begin{cases} \dfrac{\pi(x, y)}{\pi(y)} & \text{if } \pi(y) > 0 \\ 1 & \text{if } \pi(y) = 0. \end{cases}$$

Strictly, Dempster's rule does not determine $\pi_{DE}(x|y)$ when $\pi(y) = 0$, but in that case we adopt the least committal value $\pi_{DE}(x|y) = 1$. This choice has no effect on the coherence properties of Dempster's rule. Conditional possibility distributions $\pi_{DE}(\cdot|y)$, $y \in \mathcal{Y}$, are always normal.

1.6. **Renormalised Hisdal rules.** As we have seen in Section 1.2, an arbitrary solution $\pi(\cdot|y)$ of Hisdal's equation need not be normal. We can make it normal by dividing it by its highest value $\kappa(y) = \max\{\pi(x|y) \colon x \in \mathcal{X}\}$, which according to (2) is assumed in an x which maximises $\pi(\cdot, y)$, i.e. $\pi(x, y) = \pi(y)$. Consequently $\pi(y) \leq \kappa(y) \leq 1$. If we agree to treat all x which maximise $\pi(\cdot, y)$ in the same way, i.e. give them the same value of $\pi(x|y) = \kappa(y)$, then the renormalised conditional possibility distribution becomes, for $x \in \mathcal{X}$:

$$\pi_\kappa(x|y) = \begin{cases} \pi(x, y)/\kappa(y) & \text{if } \pi(x, y) < \pi(y) \\ 1 & \text{if } \pi(x, y) = \pi(y). \end{cases}$$

Given a normal joint distribution $\pi(\cdot, \cdot)$, we can therefore produce *normal* conditional distributions $\pi_\kappa(\cdot|y)$, $y \in \mathcal{Y}$, which are intermediate[3] between $\pi_{DP}(\cdot|y)$ and $\pi_{DE}(\cdot|y)$ by considering any function $\kappa \colon \mathcal{Y} \to [0, 1]$ such that $\pi(u) \leq \kappa(u) \leq 1$ for all u in $\mathcal{Y}$. This gives Dempster's rule for $\kappa(\cdot) = \pi(\cdot)$ and the Dubois-Prade rule for $\kappa = 1$. In the sense described above, these intermediate rules produce the renormalised versions of the (not necessarily normal) solutions to Hisdal's equation. They will therefore be called *renormalised Hisdal rules*.

[3]Dempster's rule yields conditional possibility distributions which dominate the ones given by the Dubois-Prade rule: $\pi_{DP}(x|y) \leq \pi_{DE}(x|y)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, and there is a strict inequality $\pi_{DP}(x|y) < \pi_{DE}(x|y)$ if and only if $0 < \pi(x, y) < \pi(y) < 1$. (In that case Dempster's rule is incompatible with Hisdal's equation.)

1.7. **Nguyen's rule.** In the rule proposed by Nguyen [17], $\pi(x|y)$ depends on $\pi(x)$ as well as on $\pi(y)$ and $\pi(x, y)$:

$$\pi_{NG}(x|y) = \begin{cases} \dfrac{\pi(x,y)}{\pi(y)} \min\{\pi(x), \pi(y)\} & \text{if } \pi(y) > 0 \\ 1 & \text{if } \pi(y) = 0. \end{cases}$$

This can be regarded as a modification of Dempster's rule: $\pi_{NG}(x|y) \leq \pi_{DE}(x|y)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, and the inequality is strict whenever $\pi(x) < 1$ or $\pi(y) < 1$. Nguyen's rule does not in general lead to solutions of Hisdal's equation. Somewhat similarly to Zadeh's, Nguyen's rule yields an unnormalised conditional possibility distribution $\pi_{NG}(\cdot|y)$ whenever $0 < \pi(y) < 1$.

1.8. **Transformed Dempster's rules.** Various authors [2, 1, 5, 11] have suggested modifying Hisdal's equation by letting a more general operator, e.g. a triangular norm, take the place of min in (1). This leads to the following relation between the conditional, marginal and joint distributions:

$$\pi(x, y) = T(\pi(x|y), \pi(y)), \quad x \in \mathcal{X},\ y \in \mathcal{Y}, \tag{3}$$

where T is a *triangular norm* [19, 20], i.e. an non-decreasing, commutative and associative binary operator on the real unit interval $[0, 1]$, with neutral element 1 and absorbing element 0. There may in general be more than one value of $\pi(x|y)$ which satisfies (3) for given $\pi(x, y)$ and $\pi(y)$. As in the special case of the Dubois-Prade conditioning rule (Section 1.4), it therefore seems reasonable to let $\pi(x|y)$ be the greatest (least informative, least specific) solution of (3). Such a solution is guaranteed to exist only if T is continuous. This yields the Dubois-Prade rule for the choice $T = \min$, and Dempster's rule if we let T be the (algebraic) product.

In recent work, De Baets *et al.* [1, 2] have argued that T should furthermore be strict, i.e. be Archimedean and have strictly increasing partial mappings. This effectively restricts T to triangular norms which are related to the (algebraic) product in the following way: $T = T_\phi$, where ϕ is an order preserving permutation of the real unit interval $[0, 1]$ (i.e. a continuous, strictly increasing function with $\phi(0) = 0$ and $\phi(1) = 1$), and

$$T_\phi(a, b) = \phi^{-1}(\phi(a)\phi(b)), \quad (a, b) \in [0, 1]^2.$$

With any such ϕ, we may therefore associate a conditioning rule as follows:

$$\pi_\phi(x|y) = \begin{cases} \phi^{-1}\big(\dfrac{\phi(\pi(x,y))}{\phi(\pi(y))}\big) & \text{if } \pi(y) > 0 \\ 1 & \text{if } \pi(y) = 0. \end{cases}$$

In an obvious way, this rule can be seen as a ϕ-transform of Dempster's rule. It always produces normal conditional possibility distributions $\pi_\phi(\cdot|y)$, $y \in \mathcal{Y}$. Dempster's rule is recovered if we let $\phi(a) = a^\theta$, $a \in [0, 1]$, where $\theta > 0$.

1.9. **Natural extension.** Another rule for conditioning possibility measures is produced by the general technique of natural extension [24]. Let co$A$ denote the set-theoretic complement of A. For $y \in \mathcal{Y}$, define $\beta(y) = \overline{P}(\mathcal{X} \times \text{co}\{y\}) = \max\{\pi(v) \colon v \in \mathcal{Y},\ v \neq y\}$. Natural extension produces the conditional possibilities:

$$\pi_{NE}(x|y) = \begin{cases} \dfrac{\pi(x,y)}{\pi(x,y) + 1 - \max\{\pi(x,y), \beta(y)\}} & \text{if } \beta(y) < 1 \\ 1 & \text{if } \beta(y) = 1. \end{cases}$$

The conditional distributions $\pi_{NE}(\cdot|y)$ are always normal. Of course $\pi_{NE}(y|x)$ is defined by analogous formulae, with $\beta(y)$ replaced by $\eta(x) = \overline{P}(\text{co}\{x\} \times \mathcal{Y}) = \max\{\pi(u) \colon u \in \mathcal{X},\ u \neq x\}$.

There are several ways of deriving the formulae for natural extension. Firstly, we will show in Section 5 that $\pi_{NE}(\cdot|y)$ is the *largest* (or least informative) conditional possibility distribution that is coherent with the joint distribution $\pi(\cdot, \cdot)$. A second derivation is based on the behavioural interpretation of $\overline{P}(A)$ as a marginally acceptable rate for betting against A, as explained more in detail in Section 2: it can be shown that $\overline{P}_{NE}(B|y) = \max\{\pi_{NE}(x|y)\colon x \in B\}$ is the lowest conditional betting rate that can be constructed by combining marginally acceptable unconditional bets:

$$\overline{P}_{NE}(B|y) = \inf\{\mu\colon I_{\mathcal{X}\times\{y\}}(\mu - I_{B\times\mathcal{Y}}) \geq \sum_{A\subseteq\mathcal{X}\times\mathcal{Y}} \lambda(A)G(A) \quad \text{for some non-negative function } \lambda \text{ defined on the power set of } \mathcal{X}\times\mathcal{Y}\},$$

or in other words, the natural extension $\overline{P}_{NE}(B|y)$ is the infimum value of μ such that the reward from a bet against B conditional on y at betting rate μ (which is the left-hand side of the inequality) is everywhere at least as large as a positive linear combination of marginally acceptable unconditional gambles. The idea here is that we can be forced to accept the gamble on the right-hand side (or a gamble arbitrarily close to it) by betting against each event A at the marginally acceptable rate (or arbitrarily close to it), and hence we should be willing to accept the gamble on the left-hand side, which is equivalent to betting against B conditional on y at rate μ. In this sense, the natural extension $\pi_{NE}(\cdot|y)$ models the conditional betting behaviour that is entailed by the joint distribution $\pi(\cdot,\cdot)$.

Thirdly, $\pi_{NE}(\cdot|y)$ can be derived as the upper envelope of all conditional probability measures $P(\cdot|y)$ which are consistent with some joint probability measure that is dominated by the possibility measure $\overline{P}$, where $\overline{P}(A) = \max\{\pi(x,y)\colon (x,y) \in A\}$. Formally, let $\mathcal{P}$ be the set of all probability measures on $\mathcal{X}\times\mathcal{Y}$, and let $\mathcal{M}$ be the set of all probability measures dominated by $\overline{P}$:

$$\mathcal{M} = \{P\colon P \in \mathcal{P},\ P(A) \leq \overline{P}(A) \text{ for every } A \subseteq \mathcal{X}\times\mathcal{Y}\}.$$

If $\beta(y) < 1$ then $P(\mathcal{X}\times\{y\}) > 0$ for each P in $\mathcal{M}$, and we may apply Bayes' rule to obtain the conditional probability $P(B|y) = P(B\times\{y\})/P(\mathcal{X}\times\{y\})$, for any $B \subseteq \mathcal{X}$. By taking the supremum over $\mathcal{M}$, we obtain the *upper envelope* of the conditional probability measures, $\overline{P}_{NE}(B|y) = \sup\{P(B|y)\colon P \in \mathcal{M}\}$, for all $B \subseteq \mathcal{X}$. If $\beta(y) = 1$ then there is P in $\mathcal{M}$ for which $P(\mathcal{X}\times\{y\}) = 0$, and *every* conditional probability measure $P(\cdot|y)$ is consistent with such a P since Bayes' rule is vacuous. The upper envelope of all these conditional probability measures is the vacuous upper probability $\overline{P}_{NE}(B|y) = 1$ if $B \subseteq \mathcal{X}$ and $B \neq \emptyset$. We will show, in the Corollary to Theorem 5, that $\overline{P}_{NE}(\cdot|y)$ is the possibility measure which has possibility distribution $\pi_{NE}(\cdot|y)$.

The conditional distributions produced by natural extension dominate those produced by all the other rules considered in this paper, and indeed they are usually vacuous since $\pi_{NE}(x|y) = 1$ unless $\beta(y) < 1$. There can be at most one $y \in \mathcal{Y}$ such that $\beta(y) < 1$, namely if y is the unique modal value (or mode) of the marginal possibility distribution $\pi(\cdot)$ of Y, i.e. the only value of v in $\mathcal{Y}$ such that $\pi(v) = 1$. If $\pi(\cdot)$ is *plurimodal*, i.e. if it has more than one mode, then $\pi_{NE}(x|y) = 1$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ and conditioning by natural extension is completely uninformative.

1.10. **Regular extension.** The process of natural extension can be modified slightly to produce conditional possibility distributions which are slightly more informative. If $\pi(y) > 0$, there are P in $\mathcal{M}$ such that $P(\mathcal{X}\times\{y\}) > 0$, and we define the *regular extension* as in [24, Appendix J]:

$$\overline{P}_{RE}(B|y) = \sup\{P(B|y)\colon P \in \mathcal{M},\ P(\mathcal{X}\times\{y\}) > 0\} \quad \text{for all } B \subseteq \mathcal{X}.$$

(The natural extension is given by the same formula in the special case where $\beta(y) < 1$.) If $\pi(y) = 0$ then $P(\mathcal{X}\times\{y\}) = 0$ for all P in $\mathcal{M}$ and we let $\overline{P}_{RE}(\cdot|y)$ be the vacuous upper probability. This process of regular extension again produces $\overline{P}_{RE}(\cdot|y)$ that are possibility measures, with possibility distributions given by:

$$\pi_{RE}(x|y) = \begin{cases} \dfrac{\pi(x,y)}{\pi(x,y) + 1 - \max\{\pi(x,y), \beta(y)\}} & \text{if } \beta(y) < 1 \\ 0 & \text{if } \beta(y) = 1 \text{ and } \pi(y) > \pi(x,y) = 0 \\ 1 & \text{otherwise.} \end{cases}$$

For possibility distributions, the natural extension is almost the same as the regular extension; they disagree only when $\beta(y) = 1$ and $\pi(y) > \pi(x,y) = 0$, in which case $\pi_{NE}(x|y) = 1$ and $\pi_{RE}(x|y) = 0$.

1.11. **Harmonic mean rule.** We will see that, of the rules defined so far, only Dempster's rule (and some of its transforms) and natural (or regular) extension are coherent in the simplest situation. However, both rules have defects: Dempster's rule produces values $\pi_{DE}(x|y)$ that are *too small* to be coherent in more complicated situations, whereas natural (or regular) extension produces values $\pi_{NE}(x|y)$ that are usually *too large* to be useful. This suggests a new rule which takes $\pi(x|y)$ to be some kind of average of $\pi_{DE}(x|y)$ and $\pi_{NE}(x|y)$. It turns out that the *harmonic mean* is a suitable average; the new values $\pi_{HM}(x|y)$ are usually non-vacuous and are coherent in both problems we study.

The harmonic mean rule produces the conditional possibilities:

$$\pi_{HM}(x|y) = \begin{cases} \dfrac{2\pi(x,y)}{\pi(x,y)+\pi(y)+1-\max\{\pi(x,y),\beta(y)\}} & \text{if } \pi(y) > 0 \\ 1 & \text{if } \pi(y) = 0. \end{cases}$$

By distinguishing cases, this can be broken down into the simpler formulae:

$$\pi_{HM}(x|y) = \begin{cases} 1 & \text{if } \pi(x,y) = \pi(y) \\ 0 & \text{if } 0 = \pi(x,y) < \pi(y) \\ \dfrac{2\pi(x,y)}{\pi(x,y)+\pi(y)} & \text{if } 0 < \pi(x,y) < \pi(y) \text{ and } \beta(y) = 1 \\ \pi(x,y) & \text{if } \beta(y) \leq \pi(x,y) < \pi(y) \\ \dfrac{2\pi(x,y)}{\pi(x,y)+2-\beta(y)} & \text{if } 0 < \pi(x,y) < \beta(y) < 1. \end{cases}$$

Of the five cases in the preceding formulae, the last two will be relatively infrequent because they can occur for at most one y in $\mathcal{Y}$ (which must satisfy $\beta(y) < 1$), and the third case is the most common non-trivial case. The formulae would not change if we defined $\pi_{HM}(x|y)$ to be the harmonic mean of $\pi_{DE}(x|y)$ and $\pi_{RE}(x|y)$, i.e. if we replaced natural extension by regular extension, as the two disagree only when $\pi_{DE}(x|y) = 0$, which implies $\pi_{HM}(x|y) = 0$ irrespective of whether natural or regular extension is used. Of course $\pi_{HM}(y|x)$ is defined by analogous formulae, with $\pi(y)$ replaced by $\pi(x)$ and $\beta(y)$ replaced by $\eta(x)$.

As expected, $\pi_{DE}(x|y) \leq \pi_{HM}(x|y) \leq \pi_{RE}(x|y) \leq \pi_{NE}(x|y)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, and it follows that $\pi_{HM}(\cdot|y)$ is always a normal possibility distribution. Unlike natural extension, the harmonic mean rule has the *non-vacuity property*: if $\pi(x,y) < \pi(y)$ then $\pi_{HM}(x|y) < 1$. Other properties of the rule are that $\pi_{HM}(x|y)$ is a continuous function of $\pi(x,y)$, $\pi(y)$ and $\beta(y)$, it is a non-decreasing function of $\pi(x,y)$ (and strictly increasing if $0 < \pi(x,y) < \pi(y)$), it is non-increasing in $\pi(y)$, and it is non-decreasing in $\beta(y)$.

2. Behavioural interpretation of possibility measures

We believe that, in order to build a useful theory of possibility, it is necessary to begin with a specific interpretation of degrees of possibility. To understand the practical meaning of possibility measures, i.e. how they should be *used* in practical reasoning, we need a behavioural interpretation that relates possibility measures to decisions and actions. Provided the joint possibility distribution $\pi(\cdot,\cdot)$ is *normal*, the corresponding possibility measure $\overline{P}$ is a *coherent* upper probability measure. It is therefore natural to adopt the behavioural interpretation of upper probabilities. Essentially, upper probabilities are marginally acceptable rates for betting *against* an event.

For $A \subseteq \mathcal{X} \times \mathcal{Y}$, let $G(A) = \overline{P}(A) - I_A = \max\{\pi(x,y) \colon (x,y) \in A\} - I_A$, where I_A denotes the indicator function of A. Then $G(A)$ is the net reward from a bet against A at betting rate $\overline{P}(A)$: the outcome is a gain of $\overline{P}(A)$ if A does not occur, and a loss of $1 - \overline{P}(A)$ if A does occur. The behavioural interpretation of the joint possibility distribution $\pi(\cdot,\cdot)$ is that the gambles $G(A)$ are marginally acceptable for all $A \subseteq \mathcal{X} \times \mathcal{Y}$, and moreover that positive linear combinations of such gambles are at least marginally acceptable[4]. (By marginally acceptable we mean that $G(A)+\varepsilon$ is acceptable for any strictly positive ε.) In other words, $\overline{P}(A)$ is an *infimum acceptable betting rate* for betting against A.

The value $\overline{P}(A)$ of a possibility measure is usually called the 'degree of possibility' or 'degree of plausibility' of A. This terminology is consistent with the behavioural interpretation. Intuitively, an event is plausible to the extent that we would not bet against it; and the less we are inclined to bet against an event, the closer to one should be the infimum acceptable betting rate. In the extreme case where an event A is fully plausible, there is no reason to bet against it at any odds and the only acceptable betting rate is $\overline{P}(A)=1$.

Next consider the behavioural interpretation of conditional possibility measures. Suppose that a conditioning rule is used to calculate a conditional degree of possibility $\overline{P}(B|y)=\max\{\pi(x|y)\colon x\in B\}$, where $B\subseteq\mathcal{X}$ and $y\in\mathcal{Y}$. This might be done to revise (update) uncertainty about the variable X after learning that $Y=y$, or to determine acceptable rates for betting against B when the bet will be called off unless $Y=y$. These two different purposes correspond to different interpretations of $\overline{P}(B|y)$ [24]. The first, called the *updating* interpretation, is that $\overline{P}(B|y)$ is the infimum acceptable rate for betting against B that we would adopt if we learned only that $Y=y$. Under that interpretation, $\overline{P}(\cdot|y)$ is the new (unconditional) possibility measure that we would use to represent our uncertainty about X after learning that $Y=y$. The second, called the *contingent* interpretation, is that $\overline{P}(B|y)$ is the infimum acceptable rate for betting against B contingent on $Y=y$, i.e. where the bet is called off unless $Y=y$. Under both interpretations, $\overline{P}(B|y)$ may be regarded as an infimum acceptable rate for betting against B *conditional on* $Y=y$.

Betting against B contingent on $Y=y$ at the marginal betting rate $\overline{P}(B|y)$ produces exactly the same outcome as betting against B after learning that $Y=y$; in both cases the net reward is $G(B|y)=I_{\mathcal{X}\times\{y\}}[\overline{P}(B|y)-I_{B\times\mathcal{Y}}]$, where the factor $I_{\mathcal{X}\times\{y\}}$ indicates that the bet is null (either called off or never undertaken) unless $Y=y$. It is therefore reasonable to require that $\overline{P}(B|y)$ should have the same numerical value under both interpretations. (This requirement is discussed in Section 6.1 of [24], where it is called the *Updating Principle*.) We will assume that the gambles $G(B|y)$ and their positive linear combinations are marginally acceptable for all $B\subseteq\mathcal{X}$ and $y\in\mathcal{Y}$, and similarly the gambles $G(C|x)=I_{\{x\}\times\mathcal{Y}}[\overline{P}(C|x)-I_{\mathcal{X}\times C}]$ are marginally acceptable for all $C\subseteq\mathcal{Y}$ and $x\in\mathcal{X}$.

3. Unnormalised possibility distributions

Under the behavioural interpretation, it is reasonable to require that the conditional possibility distributions $\pi(\cdot|y)$ and $\pi(\cdot|x)$, as well as the joint distribution $\pi(\cdot,\cdot)$, should be *normal* possibility distributions. If the joint distribution is not normal then $\overline{P}(\mathcal{X}\times\mathcal{Y})=\max\{\pi(x,y)\colon (x,y)\in\mathcal{X}\times\mathcal{Y}\}<1$, which means (under the behavioural interpretation of $\overline{P}$) that we are willing to bet against the sure event $\mathcal{X}\times\mathcal{Y}$ at a betting rate smaller than one, which is certain to result in a loss. Formally, an unnormalised possibility distribution $\pi(\cdot,\cdot)$ fails to *avoid sure loss* in the sense of Section 4. It is therefore necessary for $\pi(\cdot,\cdot)$ to be normal.

To argue that the conditional possibility distributions $\pi(\cdot|y)$ should also be normal, we need to consider their two interpretations. Under the updating interpretation, $\pi(\cdot|y)$ is the unconditional possibility distribution that we would adopt if we learned only that $Y=y$. Whatever value of Y is observed, we want the new possibility distribution to avoid sure loss, and this implies that $\pi(\cdot|y)$ must be normal for all $y\in\mathcal{Y}$.

Under the contingent interpretation, suppose that $\pi(\cdot|y)$ is not normal for some $y\in\mathcal{Y}$, so that $\overline{P}(\mathcal{X}|y)=\max\{\pi(x|y)\colon x\in\mathcal{X}\}<1$. Then a bet against the sure event $\mathcal{X}$ contingent on $Y=y$, at any betting rate μ such that $\overline{P}(\mathcal{X}|y)<\mu<1$, is acceptable. This bet has net reward $I_{\mathcal{X}\times\{y\}}[\mu-I_{\mathcal{X}\times\mathcal{Y}}]=-(1-\mu)I_{\mathcal{X}\times\{y\}}$. If $Y=y$ then the bet is certain to produce a loss of $1-\mu$, and otherwise the bet is called off. Such an unfavourable bet should not be acceptable, and again this implies that $\pi(\cdot|y)$ must be normal for all $y\in\mathcal{Y}$.

[4]Strictly, we need to assume that all gains and losses from betting are measured on a linear utility scale. The way in which possibility measures are used to make decisions in more general problems is outlined in [25].

Technical Remark 1. The preceding bet does not produce a 'sure loss' in the sense of Section 4, as it is not certain that $Y = y$. However, it does violate the slightly stronger consistency condition in [28] and the criterion of 'separate coherence' in [24, Section 6.2], each of which implies normality of $\pi(\cdot|y)$ for all $y \in \mathcal{Y}$. The criterion of avoiding sure loss in Section 4 implies normality of $\pi(\cdot,\cdot)$, and it also implies normality of $\pi(\cdot|y)$ in the special case where $\beta(y) = \overline{P}(\mathcal{X} \times \text{co}\{y\}) < 1$. The coherence criterion in Section 4 implies normality of $\pi(\cdot|y)$ under the weaker assumption that $\pi(y) > 0$. (Normality of $\pi(\cdot|y)$ is not necessary for coherence when $\pi(y) = 0$, but this case has little practical significance as values y with $\pi(y) = 0$ are 'not possible'.)

We will therefore require that the joint and conditional possibility distributions $\pi(\cdot,\cdot)$, $\pi(\cdot|y)$, and $\pi(\cdot|x)$ are always normal. This is equivalent to the requirement that each of the corresponding possibility measures $\overline{P}$, $\overline{P}(\cdot|y)$ and $\overline{P}(\cdot|x)$ is coherent, when regarded as an unconditional upper probability measure [8, 25, 26].

Of the conditioning rules surveyed in Section 1, only the rules of Zadeh and Nguyen can produce conditional possibility distributions that are unnormalised. Zadeh's rule produces an unnormalised $\pi_{ZA}(\cdot|y)$ whenever $\pi(y) < 1$. In that case, after we learn that $Y = y$, the updated possibility distribution $\pi_{ZA}(\cdot|y)$ produces a sure loss since it gives $\overline{P}(\mathcal{X}|y) < 1$. Indeed the same problem may arise directly from the conditional possibility distribution, without considering the associated possibility measure, as the following example shows.

Example 1. Let $\mathcal{X} = \{x_1, x_2\}$ and $\mathcal{Y} = \{y_1, y_2, \ldots, y_n\}$. Define $\pi(y)$ for all $y \in \mathcal{Y}$ such that $\pi(y_1) = 1$ and $\pi(y_i) < \frac{1}{2}$ if $i \neq 1$. To be specific, we take $\pi(y_i) = \frac{1}{4}$ whenever $i \neq 1$. Let $\pi(x, y) = \pi(y)$ for all $x \in \mathcal{X}$. Then Zadeh's conditioning rule gives $\pi_{ZA}(x|y) = \pi(x, y) = \pi(y)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Suppose that we learn that $Y = y_i$ where $i \neq 1$. Then the new betting rate for betting against $\{x\}$ is $\pi_{ZA}(x|y_i) = \frac{1}{4}$, and the gamble $\frac{1}{4} - I_{\{x\}}$ is marginally acceptable, for each $x \in \mathcal{X}$. But the net reward from the two bets against $\{x_1\}$ and $\{x_2\}$ is $\frac{1}{2} - I_{\mathcal{X}} = \frac{1}{2} - 1 = -\frac{1}{2}$, which is a sure loss. This arises directly from the values of $\pi_{ZA}(x|y)$, which are too small to be reasonable betting rates when $y \neq y_1$.

Nguyen's rule, which yields an unnormalised $\pi_{NG}(\cdot|y)$ whenever $0 < \pi(y) < 1$, has similar defects. Rules such as Zadeh's and Nguyen's, which may produce unnormalised conditional possibility distributions, are therefore unreasonable under a behavioural interpretation, and we will assume normality in the rest of this paper. Any such rule can be easily modified to normalise $\pi(\cdot|y)$, by dividing by the largest value $\max\{\pi(x|y) \colon x \in \mathcal{X}\}$. For example, normalising Zadeh's rule produces the modified rule $\pi(x|y) = \pi(x, y)/\pi(y)$ whenever $\pi(y) > 0$, which agrees with Dempster's rule, and we will see that this avoids the problems illustrated in the previous example.

4. General consistency criteria

We now examine whether a conditioning rule produces conditional possibilities $\pi(x|y)$ that are consistent or 'coherent' with the unconditional joint possibilities $\pi(x, y)$. The joint and conditional possibility distributions, $\pi(\cdot,\cdot)$ and $\pi(\cdot|y)$, generate upper probability measures $\overline{P}$ and $\overline{P}(\cdot|y)$ through the formulae in Section 1, so we need to consider whether $\overline{P}$ and $\overline{P}(\cdot|y)$ are coherent under a behavioural interpretation. A conditioning rule should specify $\pi(\cdot|y)$ for *every* possible value of y, so we should look at consistency of the set $\{\overline{P}(\cdot|y) \colon y \in \mathcal{Y}\}$ with $\overline{P}$. We will consider two problems.

Problem 1. Normal possibility distributions $\pi(\cdot,\cdot)$ and $\pi(\cdot|y)$ are defined for all $y \in \mathcal{Y}$. Hence $\overline{P}(A)$ and $\overline{P}(B|y)$ are defined for all $A \subseteq \mathcal{X} \times \mathcal{Y}$, $B \subseteq \mathcal{X}$ and $y \in \mathcal{Y}$, and $\overline{P}$ and $\overline{P}(\cdot|y)$ are normal possibility measures (for all y in $\mathcal{Y}$).

Problem 2. Normal possibility distributions $\pi(\cdot,\cdot)$, $\pi(\cdot|y)$ and $\pi(\cdot|x)$ are defined for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$. Hence, in addition to the quantities specified in Problem 1, $\overline{P}(C|x)$ is defined for all $C \subseteq \mathcal{Y}$ and $x \in \mathcal{X}$, and $\overline{P}(\cdot|x)$ is a normal possibility measure for all x in $\mathcal{X}$.

To motivate Problem 2, note that we can use a conditioning rule to condition on x just as well as on y, and there are many problems in which we would want to calculate both conditional distributions $\pi(\cdot|x)$ and $\pi(\cdot|y)$ from the joint distribution $\pi(\cdot,\cdot)$.

What do we mean by 'consistency'? We introduce two properties, called 'avoiding sure loss' and 'coherence', which are closely related to the behavioural interpretation of upper probability.

4.1. **Avoiding sure loss.** Recall from Section 2 that, according to the behavioural interpretation, the gambles $G(A) = \overline{P}(A) - I_A = \max\{\pi(x,y)\colon (x,y) \in A\} - I_A$ and $G(B|y) = I_{\mathcal{X}\times\{y\}}[\overline{P}(B|y) - I_{B\times\mathcal{Y}}] = I_{\mathcal{X}\times\{y\}}[\max\{\pi(x|y)\colon x \in B\} - I_{B\times\mathcal{Y}}]$ are marginally acceptable, for all $A \subseteq \mathcal{X}\times\mathcal{Y}$, $B \subseteq \mathcal{X}$ and $y \in \mathcal{Y}$. There are only finitely many gambles of the form $G(A)$ and $G(B|y)$, since $\mathcal{X}$ and $\mathcal{Y}$ are both finite. We say that $\overline{P}$ and $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ — or $\pi(\cdot,\cdot)$ and $\{\pi(\cdot|y)\colon y \in \mathcal{Y}\}$ — *avoid sure loss* when, for every positive linear combination of these marginally acceptable gambles, it is possible for the net outcome to be non-negative. In other words, $\overline{P}$ and $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ avoid sure loss if and only if there is no positive linear combination of marginally acceptable gambles which is certain to produce a net loss. Formally, for any non-negative real functions λ and μ, where $\lambda(A)$ is defined for all $A \subseteq \mathcal{X}\times\mathcal{Y}$ and $\mu(B,v)$ for all $B \subseteq \mathcal{X}$ and $v \in \mathcal{Y}$, there are $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that $\big[\sum_{A\subseteq\mathcal{X}\times\mathcal{Y}}\lambda(A)G(A)+\sum_{v\in\mathcal{Y}}\sum_{B\subseteq\mathcal{X}}\mu(B,v)G(B|v)\big](x,y) \geq 0$, i.e.

$$\sum_{A\subseteq\mathcal{X}\times\mathcal{Y}}\lambda(A)[\overline{P}(A) - I_A(x,y)] + \sum_{B\subseteq\mathcal{X}}\mu(B,y)[\overline{P}(B|y) - I_B(x)] \geq 0. \tag{4}$$

If this fails, the joint and conditional possibility distributions are inconsistent in the strong sense that they would lead us to accept gambles which are certain to produce an overall loss.

Turning now to Problem 2, we strengthen condition (4) by adding to the left-hand side an arbitrary positive linear combination of the marginally acceptable gambles $G(C|x)$, with non-negative weights $\tau(C,x)$, where $C \subseteq \mathcal{Y}$ and $x \in \mathcal{X}$. That is, we say that $\overline{P}$, $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ and $\{\overline{P}(\cdot|x)\colon x \in \mathcal{X}\}$ — or $\pi(\cdot,\cdot)$, $\{\pi(\cdot|y)\colon y \in \mathcal{Y}\}$ and $\{\pi(\cdot|x)\colon x \in \mathcal{X}\}$ — *avoid sure loss* when there is no positive linear combination of marginally acceptable gambles which is certain to produce a net loss. Formally, for any non-negative real functions λ, μ and τ, there are $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that

$$\sum_{A\subseteq\mathcal{X}\times\mathcal{Y}}\lambda(A)[\overline{P}(A) - I_A(x,y)] + \sum_{B\subseteq\mathcal{X}}\mu(B,y)[\overline{P}(B|y) - I_B(x)] + \sum_{C\subseteq\mathcal{Y}}\tau(C,x)[\overline{P}(C|x) - I_C(y)] \geq 0. \tag{5}$$

For both Problems 1 and 2, avoiding sure loss can be characterised in terms of probability measures, as follows.

Theorem 1. *With the earlier assumptions of Problems 1 and 2,*

1. $\overline{P}$ *and* $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ *avoid sure loss if and only if there are probability measures* P *(defined on the power set of* $\mathcal{X}\times\mathcal{Y}$*) and* $\{P(\cdot|y)\colon y \in \mathcal{Y}\}$ *(each defined on the power set of* $\mathcal{X}$*) such that*
   (a) $P(A) \leq \overline{P}(A) = \max\{\pi(x,y)\colon (x,y) \in A\}$ *whenever* $A \subseteq \mathcal{X}\times\mathcal{Y}$*,*
   (b) $P(B|y) \leq \overline{P}(B|y) = \max\{\pi(x|y)\colon x \in B\}$ *whenever* $B \subseteq \mathcal{X}$*,* $y \in \mathcal{Y}$*,*
   (c) *for all* $y \in \mathcal{Y}$*,* P *and* $P(\cdot|y)$ *satisfy* Bayes' rule*, i.e.* $P(B|y) = P(B\times\{y\})/P(\mathcal{X}\times\{y\})$ *whenever the denominator is non-zero, for all* $B \subseteq \mathcal{X}$*;*
2. $\overline{P}$*,* $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ *and* $\{\overline{P}(\cdot|x)\colon x \in \mathcal{X}\}$ *avoid sure loss if and only if there are probability measures* P *and* $\{P(\cdot|y)\colon y \in \mathcal{Y}\}$ *satisfying all the conditions of 1, and probability measures* $\{P(\cdot|x)\colon x \in \mathcal{X}\}$ *(each defined on the power set of* $\mathcal{Y}$*) such that*
   (d) $\overline{P}(C|x) \leq \overline{P}(C|x) = \max\{\pi(y|x)\colon y \in C\}$ *whenever* $C \subseteq \mathcal{Y}$*,* $x \in \mathcal{X}$*,*
   (e) P *and* $P(\cdot|x)$ *satisfy Bayes' rule for all* $x \in \mathcal{X}$*.*

Conditions (c) and (e) (Bayes' rule) are vacuous when $P(\mathcal{X}\times\{y\}) = 0$ or $P(\{x\}\times\mathcal{Y}) = 0$; in these cases (c) and (e) put no constraints whatsoever on the conditional probabilities $P(\cdot|y)$ and $P(\cdot|x)$. This is especially important when $\overline{P}$ is a possibility measure, since there are many joint probability measures P which satisfy (a) and have $P(\mathcal{X}\times\{y\}) = 0$ or $P(\{x\}\times\mathcal{Y}) = 0$ for most values of y or x. Indeed $\pi(x_1,y_1) = 1$ for some $(x_1,y_1) \in \mathcal{X}\times\mathcal{Y}$, so the degenerate probability mass function defined by $P(x_1,y_1) = 1$ and $P(x,y) = 0$ if $(x,y) \neq (x_1,y_1)$ will satisfy (a) and have $P(\mathcal{X}\times\{y\}) = P(\{x\}\times\mathcal{Y}) = 0$ whenever $x \neq x_1$ and $y \neq y_1$. Then conditions (c) and (e) become vacuous except when $y = y_1$ or $x = x_1$, and it is easy to find conditional probabilities $P(\cdot|y)$ and $P(\cdot|x)$ that satisfy the conditions of Theorem 1. This means that, as we will now verify, it is very easy for a conditioning rule to avoid sure loss.

Any rule for defining conditional possibilities which satisfies the simple condition (6) in the next theorem must avoid sure loss. This condition is satisfied by all the conditioning rules that we consider in this paper, and indeed it is hard to imagine anyone advocating a rule which violates the condition[5].

Theorem 2. *Suppose that the conditioning rule satisfies*

$$(\forall(x,y) \in \mathcal{X}\times\mathcal{Y})(\pi(x,y) = 1 \Rightarrow \pi(x|y) = 1) \tag{6}$$

Then $\pi(\cdot,\cdot)$ *and* $\{\pi(\cdot|y)\colon y \in \mathcal{Y}\}$ *avoid sure loss. Provided the rule satisfies* (6) *and the analogous condition with* x *and* y *interchanged,* $\pi(\cdot,\cdot)$*,* $\{\pi(\cdot|y)\colon y \in \mathcal{Y}\}$ *and* $\{\pi(\cdot|x)\colon x \in \mathcal{X}\}$ *avoid sure loss.*

The sufficient condition (6), and the analogous one with x and y interchanged, is satisfied by all the conditioning rules in Section 1, including the rules of Zadeh and Nguyen, so each of the rules avoids sure loss in both Problems 1 and 2. Avoiding sure loss is too weak to discriminate between these rules. However, a stronger consistency requirement, called 'coherence', does discriminate between the rules.

4.2. **Coherence.** Recall from Section 2 the behavioural interpretation of $G(A) = \overline{P}(A) - I_A$ and $G(B|y) = I_{\mathcal{X}\times\{y\}}[\overline{P}(B|y) - I_{B\times\mathcal{Y}}]$ as marginally acceptable gambles. By forming positive linear combinations of such gambles, we can construct new gambles which should be, at least, marginally acceptable. If there is such a positive linear combination of marginally acceptable gambles that is uniformly smaller than $G(A)$, then the overall effect of this combination of gambles is that we bet against A at a rate smaller than $\overline{P}(A)$. This is inconsistent with our interpretation of $\overline{P}(A)$ as an *infimum* acceptable betting rate, and in this case we say that the model is *incoherent*. A second type of incoherence can occur when a positive linear combination of marginally acceptable gambles is uniformly smaller than $G(B|y)$. These two types of incoherence indicate that the conditional possibility distributions $\pi(\cdot|y)$ are inconsistent with the joint distribution $\pi(\cdot,\cdot)$.

The formal definition of coherence is as follows. We say that $\overline{P}$ and $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ — or $\pi(\cdot,\cdot)$ and $\{\pi(\cdot|y)\colon y \in \mathcal{Y}\}$ — are *coherent* when they avoid sure loss and also satisfy the conditions that, for all non-negative real functions λ and μ, $A_o \subseteq \mathcal{X}\times\mathcal{Y}$, $B_o \subseteq \mathcal{X}$ and $v_o \in \mathcal{Y}$,

(i) there are $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that $\big[\sum_{A\subseteq\mathcal{X}\times\mathcal{Y}} \lambda(A)G(A) + \sum_{v\in\mathcal{Y}}\sum_{B\subseteq\mathcal{X}} \mu(B,v)G(B|v) - G(A_o)\big](x,y) \geq 0$, i.e.

$$\sum_{A\subseteq\mathcal{X}\times\mathcal{Y}} \lambda(A)[\overline{P}(A) - I_A(x,y)] + \sum_{B\subseteq\mathcal{X}} \mu(B,y)[\overline{P}(B|y) - I_B(x)] \geq \overline{P}(A_o) - I_{A_o}(x,y); \tag{7}$$

(ii) there are $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that $\big[\sum_{A\subseteq\mathcal{X}\times\mathcal{Y}} \lambda(A)G(A) + \sum_{v\in\mathcal{Y}}\sum_{B\subseteq\mathcal{X}} \mu(B,v)G(B|v) - G(B_o|v_o)\big](x,y) \geq 0$, i.e.

$$\sum_{A\subseteq\mathcal{X}\times\mathcal{Y}} \lambda(A)[\overline{P}(A) - I_A(x,y)] + \sum_{B\subseteq\mathcal{X}} \mu(B,y)[\overline{P}(B|y) - I_B(x)] \geq I_{\{v_o\}}(y)[\overline{P}(B_o|v_o) - I_{B_o}(x)]. \tag{8}$$

If condition (i) fails, the upper probability $\overline{P}(A_o)$ can be reduced, in effect, by using the information on the left-hand side of (7) which is provided by the joint and conditional possibility distributions. If (ii) fails, the conditional upper probability $\overline{P}(B_o|v_o)$ can be reduced in a similar way. In both cases the collection of joint and conditional possibility distributions is inconsistent.

Remark 1. These two types of incoherence have quite different significance in the problems studied in this paper, as we assume that the joint distribution $\pi(\cdot,\cdot)$ is given or fixed, and we want to define $\pi(\cdot|y)$ in terms of $\pi(\cdot,\cdot)$ using a conditioning rule. If condition (ii) failed, we could use it to obtain a more informative (i.e. smaller) value of $\overline{P}(B_o|v_o)$, and thus we could use it to correct or improve the conditioning rule. But if condition (i) fails, we must modify the joint distribution, which is assumed to be fixed, rather than the conditional ones. So failure of condition (i) seems in this problem to be a 'worse'

[5] Condition (6) is also a consequence of the coherence criterion defined in Section 4.2; this follows immediately from (12).

or less constructive type of incoherence than failure of (ii). Unfortunately, for the rules we consider, it is (i) that fails rather than (ii). That can be partly explained by the following remark.

Remark 2. Although the coherence conditions (i) and (ii) appear to be quite similar, condition (ii) is almost trivial in the present problem. In fact (ii) holds provided that the conditioning rule satisfies (6) and there is $y_1 \in \mathcal{Y}$ such that $y_1 \neq v_o$ and $\pi(y_1) = 1$. (To see that, take $(x, y) = (x_1, y_1)$ in (8), where x_1 is such that $\pi(x_1, y_1) = \pi(y_1)$.) So condition (ii) has force only when v_o is the *unique* element y of $\mathcal{Y}$ with $\pi(y) = 1$, that is, when $\beta(v_o) < 1$.

Technical Remark 2. In Problem 1, assuming normality of the joint and conditional possibility distributions, the preceding definitions of avoiding sure loss and coherence are essentially equivalent to those in Chapter 6 of [24] and in [28]. However, the following definition of coherence for Problem 2 is slightly weaker than the definition in Chapter 7 of [24] and that in [28]; it is equivalent to what is called 'weak coherence' in [24]. The stronger concept of coherence rules out combinations of acceptable bets whose net outcome is certain to be non-positive but not necessarily negative, i.e. not necessarily a 'sure loss'. For an example of the difference between the two definitions, involving degenerate possibility measures, see Example 7.3.5 of [24]. The stronger concept of coherence appears to be a reasonable consistency requirement, but it is more complicated than the weaker concept studied in this paper. As the conditioning rules that have been proposed previously violate even the weaker property of coherence, we will not study the stronger concept in this paper.

The preceding definition of coherence can be extended in a straightforward way to Problem 2 by adding terms $\sum_{C\subseteq\mathcal{Y}} \tau(C, x)[\overline{P}(C|x) - I_C(y)]$ to the left-hand sides of (7) and (8), to allow an arbitrary positive linear combination of marginally acceptable gambles $G(C|x)$, and adding a third condition, analogous to (ii), that requires consistency of $\overline{P}(C_o|u_o)$ with the other quantities, where $C_o \subseteq \mathcal{Y}$, $u_o \in \mathcal{X}$.

Avoiding sure loss and coherence can be verified in general by using linear programming techniques to solve a finite system of linear inequalities. Alternatively one can solve the dual linear program; see Theorem 3 below, or (more simply) Lemmas 2 and 3 of the Appendix. In Problem 1, coherence can be verified directly from Theorem 4, without using linear programming.

Coherence can be characterised in terms of probability measures, by strengthening the characterisation of avoiding sure loss given in Theorem 1. Whereas avoiding sure loss is equivalent to the existence of probability measures that are *dominated* by $\overline{P}$ and $\overline{P}(\cdot|y)$ and related by Bayes' rule, coherence is equivalent to $\overline{P}$ and $\overline{P}(\cdot|y)$ being *upper envelopes* of probability measures that are related by Bayes' rule. We say that $\overline{P}$ is the *upper envelope* of a set of probability measures $\{P_\gamma : \gamma \in \Gamma\}$ when $\overline{P}(A) = \sup\{P_\gamma(A) : \gamma \in \Gamma\}$ for all sets A in the domain of $\overline{P}$.

Theorem 3. *With the earlier assumptions of Problems 1 and 2,*

1. $\overline{P}$ *and* $\{\overline{P}(\cdot|y) : y \in \mathcal{Y}\}$ *are coherent if and only if, for every* $y \in \mathcal{Y}$ *and* $\gamma \in \Gamma$ *(where* Γ *is non-empty), there are probability measures* P_γ *(defined on the power set of* $\mathcal{X} \times \mathcal{Y}$*) and* $P_\gamma(\cdot|y)$ *(defined on the power set of* $\mathcal{X}$*) such that*
   (a) $\overline{P}$ *is the upper envelope of* $\{P_\gamma : \gamma \in \Gamma\}$,
   (b) *for each* $y \in \mathcal{Y}$, $\overline{P}(\cdot|y)$ *is the upper envelope of* $\{P_\gamma(\cdot|y) : \gamma \in \Gamma\}$,
   (c) *for each* $y \in \mathcal{Y}$ *and* $\gamma \in \Gamma$, P_γ *and* $P_\gamma(\cdot|y)$ *satisfy Bayes' rule:* $P_\gamma(B|y) = P_\gamma(B \times \{y\})/P_\gamma(\mathcal{X} \times \{y\})$ *whenever the denominator is non-zero and* $B \subseteq \mathcal{X}$*;*
2. $\overline{P}$, $\{\overline{P}(\cdot|y) : y \in \mathcal{Y}\}$ *and* $\{\overline{P}(\cdot|x) : x \in \mathcal{X}\}$ *are coherent if and only if, for every* $y \in \mathcal{Y}$, $x \in \mathcal{X}$ *and* $\gamma \in \Gamma$, *there are probability measures* P_γ, $P_\gamma(\cdot|y)$ *and* $P_\gamma(\cdot|x)$ *(defined on the power set of* $\mathcal{Y}$*) which satisfy the conditions of 1 and also*
   (d) *for each* $x \in \mathcal{X}$, $\overline{P}(\cdot|x)$ *is the upper envelope of* $\{P_\gamma(\cdot|x) : \gamma \in \Gamma\}$,
   (e) *for each* $x \in \mathcal{X}$ *and* $\gamma \in \Gamma$, P_γ *and* $P_\gamma(\cdot|x)$ *satisfy Bayes' rule:* $P_\gamma(C|x) = P_\gamma(\{x\} \times C)/P_\gamma(\{x\} \times \mathcal{Y})$ *whenever the denominator is non-zero and* $C \subseteq \mathcal{Y}$.

It follows from results in Section 6.5.7 of [24] that probability measures P and $\{P(\cdot|y) : y \in \mathcal{Y}\}$ are coherent if and only if they satisfy Bayes' rule (c). So part 1 of Theorem 3 says that $\overline{P}$ and $\{\overline{P}(\cdot|y) : y \in \mathcal{Y}\}$ are coherent if and only if they are upper envelopes of sets $\{P_\gamma : \gamma \in \Gamma\}$ and $\{P_\gamma(\cdot|y) : \gamma \in \Gamma\}$ where the corresponding probability measures P_γ and $\{P_\gamma(\cdot|y)\colon y \in \mathcal{Y}\}$ are coherent for every $\gamma \in \Gamma$. Similarly in Problem 2 it can be shown that the two versions of Bayes' rule, (c) and (e), are necessary and sufficient for coherence of probability measures P, $\{P(\cdot|y)\colon y \in \mathcal{Y}\}$ and $\{P(\cdot|x)\colon x \in \mathcal{X}\}$, so again $\overline{P}$, $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ and $\{\overline{P}(\cdot|x)\colon x \in \mathcal{X}\}$ are coherent if and only if they are upper envelopes of coherent probability measures.

All the conditioning rules we consider start with a normal joint possibility distribution $\pi(\cdot,\cdot)$ and produce conditional possibility distributions $\pi(\cdot|y)$, $y \in \mathcal{Y}$, and $\pi(\cdot|x)$, $x \in \mathcal{X}$. We say that a conditioning rule is coherent in Problem 1 if for all normal joint distributions $\pi(\cdot,\cdot)$ it yields conditional distributions $\pi(\cdot|y)$, $y \in \mathcal{Y}$, such that $\pi(\cdot,\cdot)$ and $\{\pi(\cdot|y)\colon y \in \mathcal{Y}\}$ are coherent. A similar definition applies to coherence of a conditioning rule in Problem 2.

4.3. **Consistency in terms of the possibility distributions.** In verifying condition (4), we need only to consider sets A of the form $A(x,y) = \{(u,v) \in \mathcal{X} \times \mathcal{Y}\colon \pi(u,v) \leq \pi(x,y)\}$ for some $(x,y) \in \mathcal{X} \times \mathcal{Y}$. That is because any A can be replaced by the larger set $A(x,y)$ where (x,y) maximises $\pi(u,v)$ over A, since $\overline{P}(A(x,y)) = \overline{P}(A) = \pi(x,y)$ and hence $G(A(x,y)) \leq G(A)$. Similarly we need only to consider sets B of the form $B(x|y) = \{u \in \mathcal{X}\colon \pi(u|y) \leq \pi(x|y)\}$ for some $x \in \mathcal{X}$. This observation, together with the fact that possibility measures are zero at the empty set, allows us to restate the condition for avoiding sure loss purely in terms of the joint and conditional possibility distributions: $\pi(\cdot,\cdot)$ and $\{\pi(\cdot|y)\colon y \in \mathcal{Y}\}$ avoid sure loss if and only if, for every function $\lambda\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$ and $\mu\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$ there are $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that

$$\sum_{(u,v)\in\mathcal{X}\times\mathcal{Y}} \lambda(u,v)[\pi(u,v) - \Delta_\pi(x,y;u,v)] + \sum_{w\in\mathcal{X}} \mu(w,y)[\pi(w|y) - \Delta_\pi(x;w|y)] \geq 0, \tag{9}$$

where $\mathbb{R}^+$ is the set of non-negative real numbers and

$$\Delta_\pi(x,y;u,v) = \begin{cases} 1 & \text{if } \pi(x,y) \leq \pi(u,v) \\ 0 & \text{otherwise} \end{cases}$$

$$\Delta_\pi(x;w|y) = \begin{cases} 1 & \text{if } \pi(x|y) \leq \pi(w|y) \\ 0 & \text{otherwise.} \end{cases}$$

A similar characterisation can be given for avoiding sure loss in Problem 2, with an additional term $\sum_{z\in\mathcal{Y}} \tau(x,z)[\pi(z|x) - \Delta_\pi(y;z|x)]$ on the left-hand side, where τ is any function from $\mathcal{X} \times \mathcal{Y}$ to $\mathbb{R}^+$, and

$$\Delta_\pi(y;z|x) = \begin{cases} 1 & \text{if } \pi(y|x) \leq \pi(z|x) \\ 0 & \text{otherwise.} \end{cases}$$

In (9), the quantity in the first square brackets is the reward from a bet against the event that the outcome (x,y) has degree of possibility no greater than (u,v). The quantity in the second square brackets is the reward from a similar bet made after observing y. So the left-hand side of (9) is a weighted non-negative combination of unconditional bets and bets conditional on y. The condition says that there is a possible outcome (x,y) for which the net reward is non-negative.

The same observations can be made for the coherence conditions (7) and (8). In addition, in verifying condition (7), we can restrict ourselves to sets A_o that are singletons. Any A_o can be replaced by the smaller set $\{(x_o,y_o)\}$ where (x_o,y_o) maximises $\pi(\cdot,\cdot)$ over A_o, because $\overline{P}(A_o) = \overline{P}(\{(x_o,y_o)\}) = \pi(x_o,y_o)$ and hence $G(A_o) \leq G(\{(x_o,y_o)\})$. Similarly, in (8) we need only to consider sets B_o that are singletons. Hence $\pi(\cdot,\cdot)$ and $\{\pi(\cdot|y)\colon y \in \mathcal{Y}\}$ are coherent if and only if they avoid sure loss and also, for every function $\lambda\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$ and $\mu\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$, and for every (x_o,y_o) and (w_o,z_o) in $\mathcal{X} \times \mathcal{Y}$

(i) there are $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that

$$\sum_{(u,v)\in\mathcal{X}\times\mathcal{Y}} \lambda(u,v)[\pi(u,v) - \Delta_\pi(x,y;u,v)] + \sum_{w\in\mathcal{X}} \mu(w,y)[\pi(w|y) - \Delta_\pi(x;w|y)] \geq \pi(x_o,y_o) - I_{\{(x_o,y_o)\}}(x,y); \tag{10}$$

(ii) there are $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that

$$\sum_{(u,v)\in\mathcal{X}\times\mathcal{Y}} \lambda(u,v)[\pi(u,v) - \Delta_\pi(x,y;u,v)] + \sum_{w\in\mathcal{X}} \mu(w,y)[\pi(w|y) - \Delta_\pi(x;w|y)] \geq I_{\{z_o\}}(y)[\pi(w_o|z_o) - I_{\{w_o\}}(x)]. \quad (11)$$

A similar characterisation can be given for coherence in Problem 2, only now with the additional term $\sum_{z\in\mathcal{Y}} \tau(x,z)[\pi(z|x) - \Delta_\pi(y;z|x)]$ appearing in the left-hand sides, where τ is any function from $\mathcal{X} \times \mathcal{Y}$ to $\mathbb{R}^+$, and a third condition analogous to (ii) but with the right-hand side of (11) replaced by $I_{\{u_o\}}(x)[\pi(v_o|u_o) - I_{\{v_o\}}(y)]$, where u_o can take any value in $\mathcal{X}$ and v_o any value in $\mathcal{Y}$.

In both (10) and (11), the left-hand side is the same weighted combination of bets that appears in (9). The right-hand side in (10) is the reward from a bet against (x_o, y_o). If (10) fails then $\pi(x_o, y_o)$ is too high: it can be reduced by combining the bets on the left-hand side to produce, in effect, a bet against (x_o, y_o) at a rate lower than $\pi(x_o, y_o)$. Similarly, if (11) fails then $\pi(w_o|z_o)$ is too high.

It should by now come as no surprise that the conditions of Theorem 1 can also be formulated in terms of the joint and conditional possibility distributions. Indeed, conditions (a), (b) and (d) of Theorem 1 are respectively equivalent to

(a') $P(\{(u,v) \in \mathcal{X} \times \mathcal{Y} \colon \pi(u,v) \leq \pi(x,y)\}) \leq \pi(x,y)$ whenever $x \in \mathcal{X}$, $y \in \mathcal{Y}$
(b') $P(\{u \in \mathcal{X} \colon \pi(u|y) \leq \pi(x|y)\}|y) \leq \pi(x|y)$ whenever $x \in \mathcal{X}$, $y \in \mathcal{Y}$
(d') $P(\{v \in \mathcal{Y} \colon \pi(v|x) \leq \pi(y|x)\}|x) \leq \pi(y|x)$ whenever $x \in \mathcal{X}$, $y \in \mathcal{Y}$.

5. Coherence in Problem 1

In this section we give a simple characterisation of coherence in Problem 1, where a joint possibility distribution $\pi(\cdot,\cdot)$ and conditional possibility distributions $\pi(\cdot|y)$ are specified. First we derive two necessary conditions for coherence, which will later be shown to be sufficient. Suppose that, under the assumptions of Problem 1, $\pi(\cdot,\cdot)$ and $\{\pi(\cdot|y) \colon y \in \mathcal{Y}\}$ are coherent. Then, for all x in $\mathcal{X}$ and y in $\mathcal{Y}$,

$$\pi(x|y) \geq \pi_{DE}(x|y) = \frac{\pi(x,y)}{\pi(y)} \quad \text{if } \pi(y) > 0 \quad (12)$$

and

$$\pi(x|y) \leq \pi_{NE}(x|y), \quad (13)$$

where $\pi_{NE}(x|y)$ is defined by the formulae in Section 1.9. Thus any coherent rule must produce conditional possibilities which lie between the values defined by Dempster's rule and natural extension. Since $\pi(y) \leq 1$ and $\pi(x|y) \geq 0$, it follows from (12) that $\pi(x|y) \geq \pi(x,y)$ for all x in $\mathcal{X}$ and y in $\mathcal{Y}$.

The inequalities (12) and (13) can be derived directly from (7) and (8) in the definition of coherence (see the proof of Theorem 4 in the Appendix). However, it is more insightful to consider the betting arguments that lead to these inequalities, as these demonstrate the way in which violation of (12) or (13) produces inconsistency. First consider a bet against $\{y\}$, with stake $\pi(x|y)$, at the marginally acceptable betting rate $\pi(y)$. The net reward from this bet is $\pi(x|y)[\pi(y) - I_{\mathcal{X}\times\{y\}}]$. Consider also a marginally acceptable bet against $\{x\}$ conditional on y, with stake 1, at betting rate $\pi(x|y)$. This has reward $I_{\mathcal{X}\times\{y\}}[\pi(x|y) - I_{\{x\}\times\mathcal{Y}}]$. When the two bets are combined, the overall reward is the sum $\pi(x|y)\pi(y) - I_{\mathcal{X}\times\{y\}}I_{\{x\}\times\mathcal{Y}} = \pi(x|y)\pi(y) - I_{\{(x,y)\}}$, which is the reward from a bet against $\{(x,y)\}$ at betting rate $\pi(x|y)\pi(y)$. Because $\pi(x,y)$ is interpreted as an *infimum* acceptable betting rate, it is necessary for coherence that $\pi(x,y) \leq \pi(x|y)\pi(y)$, which implies (12).

A similar argument can be used to establish (13). Consider marginally acceptable bets against $\{(x,y)\}$, with stake $1 - \pi(x|y)$ and betting rate $\pi(x,y)$, and against $\{(x,y)\} \cup \mathcal{X} \times \text{co}\{y\}$, with stake $\pi(x|y)$ and betting rate $\overline{P}(\{(x,y)\} \cup \mathcal{X} \times \text{co}\{y\}) = \max\{\pi(u,v) \colon (u,v) = (x,y) \text{ or } v \neq y\} = \max\{\pi(x,y), \beta(y)\}$. Writing $I_{\{(x,y)\}\cup\mathcal{X}\times\text{co}\{y\}} = I_{\{(x,y)\}} + 1 - I_{\mathcal{X}\times\{y\}}$ and $c = \pi(x,y) - \pi(x|y)[\pi(x,y) + 1 - \max\{\pi(x,y), \beta(y)\}]$, the overall reward from the two bets is $[1 - \pi(x|y)][\pi(x,y) - I_{\{(x,y)\}}] + \pi(x|y)[\max\{\pi(x,y), \beta(y)\} - I_{\{(x,y)\}\cup\mathcal{X}\times\text{co}\{y\}}] = I_{\mathcal{X}\times\{y\}}[\pi(x|y) - I_{\{x\}\times\mathcal{Y}}] + c$. This is the reward from a marginally acceptable bet against $\{x\}$ conditional on y, plus the constant reward c. Because $\pi(x|y)$ is interpreted as an infimum acceptable betting rate, it is necessary for coherence that $c \geq 0$, which implies (13).

Failure of (12) means that, by combining marginally acceptable gambles, we can be forced to bet against $\{(x,y)\}$ at a rate $\pi(x|y)\pi(y)$ that is lower than the asserted infimum rate $\pi(x,y)$. Similarly, failure of (13) means that we can be forced to bet against $\{x\}$ conditional on y at a rate that is lower than the asserted infimum rate $\pi(x|y)$.

The conditional possibilities produced by the Dubois-Prade rule are related to those produced by Dempster's rule and natural extension by

$$\pi_{DP}(x|y) \leq \pi_{DE}(x|y) \leq \pi_{NE}(x|y) \quad \text{for all } x \in \mathcal{X} \text{ and } y \in \mathcal{Y}.$$

It follows from (12) that the Dubois-Prade rule is incoherent whenever $\pi_{DP}(x|y) < \pi_{DE}(x|y)$, which happens whenever $0 < \pi(x,y) < \pi(y) < 1$. Formally, the Dubois-Prade rule violates coherence condition (7), taking $A_o = \{(x,y)\}$, $\lambda(A) = \pi(x,y)$ when $A = \mathcal{X} \times \{y\}$ and $\lambda(A) = 0$ otherwise, $\mu(B,v) = 1$ when $B = \{x\}$ and $v = y$, and $\mu(B,v)$ otherwise. To illustrate what goes wrong with the Dubois-Prade rule, consider a simple numerical example.

Example 2. Let the two possibility spaces be $\mathcal{X} = \{a,b\}$ and $\mathcal{Y} = \{c,d\}$. Define a joint possibility distribution $\pi(\cdot,\cdot)$ by $\pi(a,c) = \pi(b,c) = 1$, $\pi(a,d) = \frac{1}{4}$ and $\pi(b,d) = \frac{1}{5}$. Since $\pi(b,d) < \pi(d) = \frac{1}{4}$, the Dubois-Prade rule gives $\pi_{DP}(b|d) = \pi(b,d) = \frac{1}{5}$. This value is much smaller than the lower bound in (12), $\pi_{DE}(b|d) = \pi(b,d)/\pi(d) = \frac{4}{5}$, and therefore $\pi_{DP}(b|d)$ is incoherent with $\pi(\cdot,\cdot)$. If we take $\pi_{DP}(b|d) = \frac{1}{5}$ as a conditional betting rate, combining the two marginally acceptable gambles in the proof of (12) is equivalent to betting against $\{(b,d)\}$ at rate $\pi_{DP}(b|d)\pi(d) = \frac{1}{20}$, which is inconsistent with the given infimum betting rate $\pi(b,d) = \frac{1}{5}$.

In [10, 12, 13, 14], Dubois and Prade propose a definition of conditional possibilities for events A and B: $\overline{P}_{DP}(A|B) = \overline{P}(A \cap B)$ if $\overline{P}(A \cap B) < \overline{P}(B)$ and $\overline{P}_{DP}(A|B) = 1$ if $\overline{P}(A \cap B) = \overline{P}(B)$. The preceding argument shows that the quantities $\overline{P}_{DP}(A|B)$, $\overline{P}(A \cap B)$ and $\overline{P}(B)$ are incoherent whenever $0 < \overline{P}(A \cap B) < \overline{P}(B) < 1$.

So the Dubois-Prade rule is not coherent even in the simpler Problem 1. It will therefore not be coherent in Problem 2 either. It also has other defects, especially that it is *discontinuous* as $\pi(x,y) \to \pi(y)$. For example, if $\mathcal{X} = \{x_1, x_2\}$, $\pi(x_1,y) = 10^{-6}$ and $\pi(x_2,y) = (1-\varepsilon)10^{-6}$, where $\varepsilon > 0$, then $\pi_{DP}(x_1|y) = 1$ and $\pi_{DP}(x_2|y) = (1-\varepsilon)10^{-6} \to 10^{-6}$ as $\varepsilon \to 0$. This seems absurd. To be fair to Dubois and Prade, it must be pointed out that they proposed their conditioning rule in an ordinal version of possibility theory, where possibility values are not restricted to the unit interval but may belong to an arbitrary chain, and they did not consider the rule in a numerical context. In the ordinal version, discontinuity and incoherence are not applicable. Nevertheless, our arguments indicate that the Dubois-Prade rule should not be used in the numerical version of possibility theory.

Similarly we can see that the renormalised versions of all the solutions to Hisdal's equation are incoherent, except for the extreme case of Dempster's rule. The argument showing incoherence of the Dubois-Prade rule shows that a renormalised Hisdal rule is incoherent whenever $\pi_\kappa(x|y) < \pi(x,y)/\pi(y)$ for some (x,y), which happens whenever there are (x,y) such that $0 < \pi(x,y) < \pi(y) < \kappa(y) \leq 1$.

An argument similar to that for the Dubois-Prade rule shows that Ramer's rule is incoherent in Problem 1 (and *a fortiori* in Problem 2). For any $y \in \mathcal{Y}$, it yields values $\pi_{RA}(x|y)$ which are incoherent with $\pi(\cdot,\cdot)$ whenever $0 < \pi(x,y) < \pi(y) < 1$. It will also be incoherent if $0 < \pi(x,y) = \pi(y) < 1$, provided that there is another x_o in $\mathcal{X}$ such that $\pi(x_o,y) = \pi(y)$ and we have assigned the value 1 to $\pi_{RA}(x_o|y)$ and not to $\pi_{RA}(x|y)$.

The conditioning rules of Zadeh and Nguyen, which can produce unnormalised conditional distributions but which do avoid sure loss in general, are also incoherent in Problem 1. Zadeh's rule violates condition (12) whenever $\pi(x,y) > 0$ and $\pi(y) < 1$. Nguyen's rule violates (12) whenever $\pi(x,y) > 0$ and $\min\{\pi(x), \pi(y)\} < 1$.

We have shown that conditions (12) and (13) are necessary for coherence. The next theorem shows that these conditions are also sufficient for coherence in Problem 1. This gives a complete characterisation of coherence in Problem 1: a conditioning rule is coherent if and only if it is intermediate between Dempster's rule and natural extension. Given a joint possibility distribution for which $\pi(y) > 0$, Dempster's rule produces the smallest (most informative) conditional possibilities that are coherent, and natural extension produces the largest (least informative) conditional possibilities that are coherent.

Theorem 4. *Normal possibility distributions $\pi(\cdot,\cdot)$ and $\{\pi(\cdot|y)\colon y \in \mathcal{Y}\}$ are coherent if and only if the values $\pi(x|y)$ satisfy*

$$\pi_{DE}(x|y) \le \pi(x|y) \le \pi_{NE}(x|y) \quad \text{if } \pi(y) > 0$$

for all x in $\mathcal{X}$ and y in $\mathcal{Y}$.

Remark 3. It is remarkable that the coherence conditions (7) and (8), which appear to impose complicated relationships between the possibility distributions $\pi(\cdot,\cdot)$ and $\{\pi(\cdot|y)\colon y \in \mathcal{Y}\}$, are actually equivalent to the simple bounds (12) and (13). Since the possible (coherent) values of $\pi(x|y)$ depend only on $\pi(x,y)$, $\pi(y)$ and $\beta(y)$, the values of $\pi(x|y)$ and $\pi(u|v)$ can be chosen independently when $(u,v) \neq (x,y)$; coherence does not impose any relationship between them, except through $\pi(\cdot,\cdot)$. These simple results do not extend to Problem 2, where the coherence relationships, e.g. (15) and (16) below, are considerably more complicated.

Remark 4. The conditional possibility $\pi(x|y)$ is uniquely determined by the inequalities in Theorem 4 if and only if $\pi(y) > 0$ and $\pi_{DE}(x|y) = \pi_{NE}(x|y)$. That happens if and only if (a) $\pi(x,y) = \pi(y) > 0$, which implies that $\pi(x|y) = 1$, or (b) $\pi(x,y) \ge \beta(y)$, which gives $\pi(x|y) = \pi(x,y)$. In other cases the interval of coherent values for $\pi(x|y)$ may be very wide. For example, if $\pi(x,y) = 0$ and $\beta(y) = 1$, or if $\pi(y) = 0$, then $\pi(x|y)$ can take any value in the interval $[0,1]$. There can be at most one value of y with $\beta(y) < 1$, and the upper bound in Theorem 4 is *vacuous*, i.e. $\pi_{NE}(x|y) = 1$, for all other y in $\mathcal{Y}$ and all x in $\mathcal{X}$.

Remark 5. When $\pi(y) = 0$, Theorem 4 imposes no restrictions on $\pi(x|y)$, and $\pi(\cdot|y)$ can be taken to be any normal possibility distribution without affecting coherence. This indeterminacy mirrors the indeterminacy of conditional probabilities $P(\cdot|y)$ in classical probability theory when $P(\mathcal{X} \times \{y\}) = 0$. The case $\pi(y) = 0$ appears to have no practical importance, at least when $\mathcal{Y}$ is finite, because the event $C = \{y \in \mathcal{Y}\colon \pi(y) = 0\}$ is 'practically impossible' in the sense that we should be willing to bet against its occurrence at any odds, since $\overline{P}(C) = 0$. But if C does not occur then the conditional distributions $\pi(\cdot|y)$ for which $\pi(y) = 0$ will have no effect on behaviour; the corresponding bets will always be called off. Notice also that if $\pi(y) = 0$ then $\pi(x,y) = 0$ for all $x \in \mathcal{X}$ and there is no useful information in the joint possibility distribution $\pi(\cdot,\cdot)$ from which $\pi(\cdot|y)$ can be constructed. It is therefore not surprising that $\pi(\cdot|y)$ can be chosen arbitrarily.

Remark 6. Theorem 4 establishes that Dempster's rule is coherent in Problem 1. However, this result relies on $\overline{P}$ being a *possibility measure*. It is well known that if $\overline{P}$ is a more general type of upper probability measure (such as a plausibility function) and $\overline{P}(\cdot|y)$ is defined by Dempster's rule $\overline{P}(B|y) = \overline{P}(B \times \{y\})/\overline{P}(\mathcal{X} \times \{y\})$, then $\overline{P}$ and $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ may be incoherent, and indeed they may not even avoid sure loss. See [25] and Section 5.13 of [24] for examples.

Theorem 4 tells us that the transformed Dempster's rule for a given order-preserving permutation ϕ of the unit interval (Section 1.8) is coherent in Problem 1 if and only if ϕ satisfies $\phi(ab) \ge \phi(a)\phi(b)$ for all $(a,b) \in [0,1]^2$, that is, if and only if the triangular norm T_ϕ is pointwise dominated by the algebraic product.

The new rule defined in Section 1.11 yields values $\pi_{HM}(x|y)$ which are the harmonic mean of the values $\pi_{DE}(x|y)$ and $\pi_{NE}(x|y)$. The new rule is therefore intermediate between Dempster's rule and natural extension, and hence coherent in Problem 1, by Theorem 4. The same is true of regular extension.

6. Coherence in Problem 2

Of the conditioning rules surveyed in the Introduction, essentially only four are coherent in Problem 1: Dempster's rule (and some of its transforms), natural extension, regular extension and the harmonic mean rule. In this section we investigate which of these rules are coherent also in Problem 2, when they are used to define both $\{\pi(\cdot|x)\colon x\in\mathcal{X}\}$ and $\{\pi(\cdot|y)\colon y\in\mathcal{Y}\}$. Since coherence is a stronger requirement in Problem 2 than in Problem 1, conditions (12) and (13), and their analogous versions with x and y interchanged, are necessary for coherence in Problem 2. The following three inequalities, which relate the quantities $\pi(x|y)$, $\pi(y|x)$ and $\pi(x,y)$, are also necessary for coherence in Problem 2:

$$\pi(x,y)\leq\frac{\pi(x|y)\pi(y|x)\max\{\pi(x),\pi(y)\}}{\pi(x|y)+\pi(y|x)-\pi(x|y)\pi(y|x)},\tag{14}$$

where $0/0$ is taken to be 0;

$$\pi(x,y)\leq\frac{\pi(x|y)\pi(y|x)\max\{\pi(y|u)\colon u\in\mathrm{co}\{x\}\}}{\pi(x|y)\max\{\pi(y|u)\colon u\in\mathrm{co}\{x\}\}+\pi(y|x)-\pi(x|y)\pi(y|x)},\tag{15}$$

provided the denominator is non-zero; and

$$\pi(x,y)\leq\frac{\pi(y|x)\pi(x|y)\max\{\pi(x|v)\colon v\in\mathrm{co}\{y\}\}}{\pi(y|x)\max\{\pi(x|v)\colon v\in\mathrm{co}\{y\}\}+\pi(x|y)-\pi(y|x)\pi(x|y)},\tag{16}$$

provided the denominator is non-zero. Conditions (14)–(16) are considerably more complicated than (12)–(13), which indicates that the coherence relationships are much more involved in Problem 2 than in Problem 1. (16) is obtained from (15) by interchanging x and y. Although (15) and (16) look similar to (14), the three conditions are logically independent.

(14) is symmetric between x and y, and it can be written more simply as

$$\frac{1}{\pi(y|x)}+\frac{1}{\pi(x|y)}\leq 1+\frac{\max\{\pi(x),\pi(y)\}}{\pi(x,y)}.$$

The three conditions generalise from possibility distributions to coherent upper probabilities, and (14) also generalises from variables to events: if A and B are any two events, it is necessary for coherence of the quantities $\overline{P}(A|B)$, $\overline{P}(B|A)$, $\overline{P}(A\cap B)$ and $\overline{P}(A\cup B)$ that

$$\overline{P}(A\cap B)\leq\frac{\overline{P}(A|B)\overline{P}(B|A)\overline{P}(A\cup B)}{\overline{P}(A|B)+\overline{P}(B|A)-\overline{P}(A|B)\overline{P}(B|A)}.\tag{17}$$

When the joint upper probability $\overline{P}$ is a probability measure and the conditional probabilities are defined by Bayes' rule, (17) is satisfied with equality.

Again we will outline the betting arguments that lead to (14) and (15). (14) holds if $\pi(x|y)=0$, since then $\pi(x,y)\leq\pi(x|y)\pi(y)=0$ by (12), and similarly if $\pi(y|x)=0$. Suppose that $\pi(x|y)>0$ and $\pi(y|x)>0$. Define $\zeta=\pi(x|y)+\pi(y|x)-\pi(x|y)\pi(y|x)$ (so $\zeta>0$), $\kappa=\pi(y|x)/\zeta$, $\nu=\pi(x|y)/\zeta$, and $\mu=\pi(x|y)\pi(y|x)/\zeta$. Consider three marginally acceptable gambles: a bet against $\{x\}$ conditional on y at rate $\pi(x|y)$ with stake κ; a bet against $\{y\}$ conditional on x at rate $\pi(y|x)$ with stake ν; and a bet against $C=(\{x\}\times\mathcal{Y})\cup(\mathcal{X}\times\{y\})$ at rate $\overline{P}(C)=\max\{\pi(x),\pi(y)\}$, with stake μ. Using the identities $I_C=I_{\{x\}\times\mathcal{Y}}+I_{\mathcal{X}\times\{y\}}-I_{\{(x,y)\}}$ and $\kappa+\nu-\mu=1$, the overall reward from the three bets is $\kappa I_{\mathcal{X}\times\{y\}}[\pi(x|y)-I_{\{x\}\times\mathcal{Y}}]+\nu I_{\{x\}\times\mathcal{Y}}[\pi(y|x)-I_{\mathcal{X}\times\{y\}}]+\mu[\overline{P}(C)-I_C]=\mu\overline{P}(C)-I_{\{(x,y)\}}$. This is the reward from a bet against $\{(x,y)\}$ at rate $\overline{P}(C)=\mu\max\{\pi(x),\pi(y)\}$. Since $\pi(x,y)$ is the infimum acceptable rate for betting against $\{(x,y)\}$, it is necessary for coherence that $\pi(x,y)\leq\mu\max\{\pi(x),\pi(y)\}$, which establishes (14). To derive (17), replace $\{x\}\times\mathcal{Y}$, $\mathcal{X}\times\{y\}$, $\pi(x|y)$, $\pi(y|x)$, $\pi(x,y)$ and $\overline{P}(C)$ in the preceding argument by A, B, $\overline{P}(A|B)$, $\overline{P}(B|A)$, $\overline{P}(A\cap B)$ and $\overline{P}(A\cup B)$ respectively.

(15) can be derived in a similar way by combining marginally acceptable bets against $\{x\}$ conditional on y, against $\{y\}$ conditional on x, and against $\{y\}$ conditional on u for every u in $\mathrm{co}\{x\}$, with stakes proportional to $\pi(y|x)$, $\pi(x|y)\max\{\pi(y|u)\colon u\in\mathrm{co}\{x\}\}$, and $\pi(x|y)\pi(y|x)$ respectively. The derivation of (16) is similar, with x and y interchanged.

The inequalities (12)–(16), and the versions of (12) and (13) with x and y interchanged, are therefore necessary for coherence in Problem 2. It appears that, in the simplest non-trivial case where the possibility spaces $\mathcal{X}$ and $\mathcal{Y}$ each have only two elements, these conditions are also sufficient for coherence.

Dempster's rule is coherent in Problem 1, when it is used to define $\pi(\cdot|y)$, but it is *not* coherent in Problem 2, when it is used to define both $\pi(\cdot|y)$ and $\pi(\cdot|x)$. To show what goes wrong with Dempster's rule in Problem 2, we give a simple numerical example.

Example 3. Let the two possibility spaces be $\mathcal{X} = \{a, b\}$ and $\mathcal{Y} = \{c, d\}$. Define the joint possibility distribution $\pi(\cdot, \cdot)$ by $\pi(b, d) = \frac{1}{2}$ and $\pi(x, y) = 1$ if $(x, y) \neq (b, d)$. Dempster's rule gives conditional possibilities $\pi_{DE}(b|d) = \pi_{DE}(d|b) = \frac{1}{2}$. These values violate (14), taking $x = b$ and $y = d$, as the right-hand side of (14) is seen to be $\frac{1}{3}$ which is smaller than $\pi(b, d) = \frac{1}{2}$. In betting terms, when we combine marginally acceptable bets against $\{b\}$ conditional on d and against $\{d\}$ conditional on b (each with stake $\frac{2}{3}$), the overall reward is no greater than that from a bet against $\{(b, d)\}$ at rate $\frac{1}{3}$. This is inconsistent with the interpretation of $\pi(b, d) = \frac{1}{2}$ as an infimum acceptable rate for betting against $\{(b, d)\}$. Conditions (15) and (16) are also violated for $(x, y) = (b, d)$.

The inconsistency can be understood in another way, using the comparative probability relation 'is at least as probable as', which we denote by $\succeq$. (That is possible because the three values $\pi(b, d)$, $\pi_{DE}(b|d)$ and $\pi_{DE}(d|b)$ in the preceding argument are each $\frac{1}{2}$.) When $\pi(\omega) = \frac{1}{2}$ we can infer that $\mathrm{co}\{\omega\} \succeq \{\omega\}$ (i.e. $\{\omega\}$ is no more probable than its complement), since $\underline{P}(\mathrm{co}\{\omega\}) = 1 - \pi(\omega) = \frac{1}{2} \geq \pi(\omega) = \overline{P}(\{\omega\})$. Now suppose we learn that $x = b$. This reduces the joint possibility space to $\{(b, c), (b, d)\}$. Since $\pi_{DE}(d|b) = \frac{1}{2}$, $c \succeq d$ after learning $x = b$. Because the only effect of learning $x = b$ is to eliminate two possibilities, we must have $(b, c) \succeq (b, d)$ initially. Similarly, because $\pi_{DE}(b|d) = \frac{1}{2}$, $(a, d) \succeq (b, d)$ initially. Thus, initially, both (b, c) and (a, d) are at least as probable as (b, d). The maximum possible upper probability for (b, d) is therefore $\frac{1}{3}$, which is inconsistent with the given upper probability $\pi(b, d) = \frac{1}{2}$.

Dempster's rule produces incoherence quite generally in Problem 2. By substituting the values $\pi_{DE}(x|y) = \pi(x, y)/\pi(y)$ and $\pi_{DE}(y|x) = \pi(x, y)/\pi(x)$ in (14), we find that Dempster's rule satisfies (14) only when $\pi(x, y) = 0$ or $\pi(x, y) = \min\{\pi(x), \pi(y)\}$. (It can be shown, using Theorem 3, that Dempster's rule defines coherent conditional possibilities in Problem 2 if and only if, for every (x, y) in $\mathcal{X} \times \mathcal{Y}$, either $\pi(x, y) = 0$ or $\pi(x, y) = \min\{\pi(x), \pi(y)\}$.) Thus Dempster's rule produces incoherence in Problem 2 whenever there are (x, y) in $\mathcal{X} \times \mathcal{Y}$ such that $0 < \pi(x, y) < \min\{\pi(x), \pi(y)\}$. In that case the values $\pi_{DE}(x|y)$ and $\pi_{DE}(y|x)$ are *too small* to achieve coherence with $\pi(\cdot, \cdot)$.

In the following example, we show that any transformed Dempster's rule may produce values $\pi_\phi(\cdot|\cdot)$ that are incoherent in Problem 2 (irrespective of the choice of ϕ).

Example 4. Let ϕ be an order preserving permutation of $[0, 1]$. Let the two possibility spaces be $\mathcal{X} = \{a, b\}$ and $\mathcal{Y} = \{c, d\}$. As in Example 3, define the joint possibility distribution $\pi(\cdot, \cdot)$ by $\pi(b, d) = \frac{1}{2}$ and $\pi(x, y) = 1$ if $(x, y) \neq (b, d)$. For the marginal distributions, we have that $\pi(a) = \pi(b) = \pi(c) = \pi(d) = 1$, so for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$: $\pi_\phi(x|y) = \pi_\phi(y|x) = \pi(x, y) = \pi_{DE}(x|y) = \pi_{DE}(y|x)$. The rule produces the same values for the conditional possibilities as Dempster's rule, and will therefore produce incoherent values of $\pi_\phi(x|y)$ and $\pi_\phi(y|x)$ for $(x, y) = (b, d)$ (since $0 < \frac{1}{2} < 1$).

The following result establishes that natural extension produces coherent conditional possibilities in Problem 2. (In fact the general procedure of natural extension always preserves coherence, provided only that the underlying possibility spaces $\mathcal{X}$ and $\mathcal{Y}$ are finite; see [24, Section 8.1].) A slight modification of the proof shows that regular extension is also coherent.

Theorem 5. *Let $\pi(\cdot, \cdot)$ be a normal possibility distribution on $\mathcal{X} \times \mathcal{Y}$, and let $\pi_{NE}(\cdot|\cdot)$ denote the conditional possibility distributions produced by natural extension (Section 1.9). Then $\pi(\cdot, \cdot)$, $\{\pi_{NE}(\cdot|y) : y \in \mathcal{Y}\}$ and $\{\pi_{NE}(\cdot|x) : x \in \mathcal{X}\}$ are coherent.*

Theorems 4 and 5 together imply that $\pi_{NE}(\cdot|y)$ and $\pi_{NE}(\cdot|x)$ are the *largest* (i.e. the least informative) conditional possibility distributions which achieve coherence in Problem 2. (This is a general property of natural extension; see [24].) The following corollary gives another characterisation of $\pi_{NE}(\cdot|y)$ as the upper envelope of a set of conditional probability measures.

Corollary 1. *With the notation of Theorem 5, let $\overline{P}$ denote the possibility measure that is generated by $\pi(\cdot, \cdot)$. Then, for every y in $\mathcal{Y}$, $\pi_{NE}(\cdot|y)$ is the upper envelope of the set of all conditional probability measures $P(\cdot|y)$ that are coherent with any probability measure P that is dominated by $\overline{P}$. (Here $P(\cdot|y)$ and P are coherent if and only if they satisfy Bayes' rule, i.e. $P(B|y) = P(B \times \{y\})/P(\mathcal{X} \times \{y\})$ whenever the denominator is non-zero and $B \subseteq \mathcal{X}$.)*

Theorem 5 establishes that there are conditioning rules which are coherent in Problem 2. However, as discussed in Section 1.9, conditioning by natural extension is 'almost uninformative' in the sense that $\pi_{NE}(\cdot|y)$ can be non-vacuous for at most one value of y. It is desirable to find a conditioning rule that is coherent in Problem 2 and typically produces non-vacuous conditional possibility distributions. By Theorem 4, any coherent rule must satisfy: if $\pi(x,y) = \pi(y)$ then $\pi(x|y) = 1$. The strongest non-vacuity property that is compatible with coherence is the condition:

$$\pi(x,y) < \pi(y) \Rightarrow \pi(x|y) < 1. \tag{18}$$

It turns out that the harmonic mean rule, which has this property, is coherent in Problem 2.

Theorem 6. *Let $\pi(\cdot,\cdot)$ be a normal possibility distribution on $\mathcal{X} \times \mathcal{Y}$, and let $\pi_{HM}(\cdot|\cdot)$ denote the conditional possibility distributions produced by the harmonic mean rule (defined in Section 1.11). Then $\pi(\cdot,\cdot)$, $\{\pi_{HM}(\cdot|y) \colon y \in \mathcal{Y}\}$ and $\{\pi_{HM}(\cdot|x) \colon x \in \mathcal{X}\}$ are coherent.*

In the numerical examples that we have studied, the conditional possibilities $\pi_{HM}(x|y)$ and $\pi_{HM}(y|x)$ either achieve equality or nearly do so in at least one of the inequalities (14)–(16) for most values of (x,y). By substituting the formulae for $\pi_{HM}(x|y)$ in (14), we find that the harmonic mean rule achieves equality in (14) whenever any of the following conditions is satisfied: (i) $\pi(x,y) = 1$; (ii) $\pi(x,y) = 0$; (iii) $\pi(x,y) = \pi(x) > \beta(y)$; (iv) $\pi(x,y) = \pi(y) > \eta(x)$; (v) $\pi(x) = \beta(y) < 1$; (vi) $\pi(y) = \eta(x) < 1$; or (vii) $\beta(y) = \eta(x) = 1$ and $\pi(x) = \pi(y)$. When $\mathcal{X}$ and $\mathcal{Y}$ each contain two elements, for example, there must be equality in (14) for at least two of the four (x,y) pairs, and if $\pi(x_1,y_2) = \pi(x_2,y_1)$ then there must be equality in (14) for all (x,y). For instance, in the numerical Example 3 that was given to illustrate the incoherence of Dempster's rule, the harmonic mean rule produces $\pi_{HM}(b|d) = \pi_{HM}(d|b) = \frac{2}{3}$, which achieves equality in (14), (15) and (16). (It is the smallest value that satisfies these inequalities.) Thus the harmonic mean rule appears to produce conditional possibilities that are about as small as possible (i.e. as informative as possible) to achieve coherence.

7. Conclusion

We have investigated the coherence properties of various rules for defining conditional possibilities. These properties are summarised in Table 1. The conditional possibility distributions produced by the rules of Zadeh and Nguyen are usually not normal, and they are not reasonable under the behavioural interpretation adopted in this paper. The Dubois-Prade rule, Ramer's rule and the renormalised Hisdal rules avoid sure loss, but they are typically incoherent even in the simpler Problem 1, and they therefore seem unsatisfactory under a behavioural interpretation. Dempster's rule is coherent in Problem 1 but typically not in Problem 2, when it is used to condition on x as well as on y.

Conditioning Rule	**Normality of Conditional Distributions**	**Coherence in Problem 1**	**Coherence in Problem 2**
Zadeh's rule	no	no	no
Nguyen's rule	no	no	no
Ramer's rule	yes	no	no
Dubois-Prade rule	yes	no	no
Dempster's rule	yes	yes	no
Transformed Dempster's rule	yes	depends on ϕ	no
Renormalised Hisdal rules	yes	only Dempster's rule	no
Natural extension	yes	yes	yes
Regular extension	yes	yes	yes
Harmonic mean rule	yes	yes	yes

Table 1. Coherence properties of the conditioning rules surveyed in this paper. All the rules in the table avoid sure loss in both Problems 1 and 2.

Of the rules examined in this paper, only three are coherent in both problems: natural extension, regular extension and the new harmonic mean rule. Natural extension models the information that is contained in the unconditional possibility distribution $\pi(\cdot,\cdot)$ concerning conditional possibilities. That is, the conditional betting rates $\pi_{NE}(x|y)$ that are produced by natural extension are the rates that can be constructed, by combining bets, from the unconditional betting rates alone, without providing any additional information. If the aim is to model what can be learned about conditional possibilities from the unconditional possibility distribution alone, then natural extension is the correct conditioning rule. By Theorem 4, any coherent conditional possibility distribution $\pi(\cdot|y)$ must provide at least as much information about conditional betting rates as natural extension. From this point of view, any other coherent conditioning rule is implicitly adding information to that contained in the unconditional possibility distribution.

Unfortunately the natural extension $\pi_{NE}(\cdot|y)$ may be too uninformative to be useful, as $\pi_{NE}(\cdot|y)$ can be non-vacuous for at most one possible value of y. This simply reflects the fact that the unconditional possibility distribution $\pi(\cdot,\cdot)$ alone supplies very little information, and sometimes none at all, about conditional possibilities. Regular extension is only slightly more informative than natural extension.

If it is necessary to define non-vacuous conditional possibilities then either additional information must be supplied or another conditioning rule must be used. Of the alternative rules studied in this paper, the new harmonic mean rule is the most satisfactory. It has the following desirable properties.

(a) It is coherent in both Problems 1 and 2.
(b) It generally produces non-vacuous conditional possibility distributions, in the sense that $\pi_{HM}(x|y) < 1$ whenever $\pi(x,y) < \pi(y)$.
(c) Conditional possibilities are intermediate between those produced by natural extension and Dempster's rule.
(d) The conditional distributions seem to be about as informative as possible to achieve coherence in Problem 2.
(e) It has the continuity and monotonicity properties noted in Section 1.11.

Acknowledgements

Gert de Cooman is a Postdoctoral Fellow of the Fund for Scientific Research, Flanders (FWO). He would like to thank the FWO for partially funding the research reported in this paper.

Appendix: Proofs of Results

Let $\mathcal{L}$ denote the set of all functions $Z\colon \mathcal{X}\times\mathcal{Y}\to\mathbb{R}$. Any probability measure P, defined on the power set of $\mathcal{X}\times\mathcal{Y}$, has a unique extension to a linear prevision (i.e. a positive linear functional with unit norm) on $\mathcal{L}$, defined by $P(Z) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(\{(x,y)\})Z(x,y)$ for all $Z\in\mathcal{L}$.

Lemma 1. *Let $\mathcal{D}$ be an arbitrary subset of $\mathcal{L}$. Then the following conditions are equivalent.*

(I) *There is a probability measure P (defined on the power set of $\mathcal{X}\times\mathcal{Y}$) whose associated linear prevision satisfies $P(Z)\geq 0$ for all $Z\in\mathcal{D}$.*
(II) *For any non-negative integer n and Z_1, Z_2, ..., Z_n in $\mathcal{D}$, there are $x\in\mathcal{X}$ and $y\in\mathcal{Y}$ such that $\sum_{j=1}^{n} Z_j(x,y)\geq 0$.*

Proof. This follows from Lemma 3.3.2 of [24], using the assumption that $\mathcal{X}\times\mathcal{Y}$ is finite. □

Proof of Theorem 1. Let $\mathcal{D} = \{\lambda G(A)\colon \lambda\geq 0, A\subseteq\mathcal{X}\times\mathcal{Y}\}\cup\{\mu G(B|y)\colon \mu\geq 0, B\subseteq\mathcal{X}, y\in\mathcal{Y}\}$. By condition (4), $\overline{P}$ and $\{\overline{P}(\cdot|y)\colon y\in\mathcal{Y}\}$ avoid sure loss if and only if $\mathcal{D}$ satisfies condition (II) of Lemma 1. This is equivalent to condition (I) of Lemma 1, and hence to the existence of a probability measure P which satisfies:

(i) $P(G(A)) = \overline{P}(A) - P(A)\geq 0$, i.e. $P(A)\leq\overline{P}(A)$, whenever $A\subseteq\mathcal{X}\times\mathcal{Y}$; and
(ii) $P(G(B|y)) = P(\mathcal{X}\times\{y\})\overline{P}(B|y) - P(B\times\{y\})\geq 0$, i.e. if $P(\mathcal{X}\times\{y\})>0$ then $P(B\times\{y\})/P(\mathcal{X}\times\{y\})\leq\overline{P}(B|y)$, for all $B\subseteq\mathcal{X}$ and $y\in\mathcal{Y}$.

Thus avoiding sure loss is equivalent to the existence of a P which satisfies (i) and (ii). It remains to be proven that this is equivalent to the existence of probability measures which satisfy (a)–(c) of Theorem 1. If probability measures P and $\{P(\cdot|y)\colon y \in \mathcal{Y}\}$ satisfy (a)–(c) then P satisfies (i), which is identical to (a), and (ii), which follows from (b) and (c). (Condition (ii) holds when $P(\mathcal{X} \times \{y\}) = 0$ since then $P(G(B|y)) = 0$.) Conversely, suppose that a probability measure P satisfies (i) and (ii). Then P satisfies (a) of the Theorem. Define conditional probability measures $P(\cdot|y)$ by $P(B|y) = P(B \times \{y\})/P(\mathcal{X} \times \{y\})$ whenever $P(\mathcal{X} \times \{y\}) > 0$, so that (c) is satisfied. Then (b) holds whenever $P(\mathcal{X} \times \{y\}) > 0$, by (ii). Because each $\overline{P}(\cdot|y)$ is a normal possibility measure, it is a coherent upper probability measure, and it follows from Theorem 3.3.3 of [24] that, for each $y \in \mathcal{Y}$, there is a probability measure $P(\cdot|y)$ which satisfies (b). If $P(\mathcal{X} \times \{y\}) = 0$ then $P(\cdot|y)$ can be taken to be any such probability measure. Thus (a)–(c) are satisfied. This completes the proof of part 1 of the Theorem. Part 2 can be proven by a similar argument, taking $\mathcal{D} = \{\lambda G(A)\colon \lambda \geq 0, A \subseteq \mathcal{X} \times \mathcal{Y}\} \cup \{\mu G(B|y)\colon \mu \geq 0, B \subseteq \mathcal{X}, y \in \mathcal{Y}\} \cup \{\tau G(C|x)\colon \tau \geq 0, C \subseteq \mathcal{Y}, x \in \mathcal{X}\}$, and using Lemma 1 to show that avoiding sure loss is equivalent to the existence of a probability measure P which satisfies (i), (ii) and

(iii) if $P(\{x\} \times \mathcal{Y}) > 0$ then $P(\{x\} \times C)/P(\{x\} \times \mathcal{Y}) \leq \overline{P}(C|x)$, for all $C \subseteq \mathcal{Y}$ and $x \in \mathcal{X}$.

By the preceding argument, (i) and (ii) are equivalent to (a)–(c), and an analogous argument shows that (iii) is equivalent to (d) and (e). □

Proof of Theorem 2. First consider Problem 1. Since $\pi(\cdot,\cdot)$ is normal, there are $x_1 \in \mathcal{X}$ and $y_1 \in \mathcal{Y}$ such that $\pi(x_1, y_1) = 1$. To verify (4), take $x = x_1$ and $y = y_1$ in (4). If $(x_1, y_1) \in A$ then $\overline{P}(A) - I_A(x_1, y_1) = 1 - 1 = 0$. If $x_1 \in B$ then $\overline{P}(B|y_1) - I_B(x_1) = 1 - 1 = 0$, since $\overline{P}(B|y_1) \geq \pi(x_1|y_1) = 1$ by (6). Hence the left hand side of (4) simplifies to

$$\sum_{A \subseteq \mathcal{X} \times \mathcal{Y}, (x_1, y_1) \notin A} \lambda(A)\overline{P}(A) + \sum_{B \subseteq \mathcal{X}, x_1 \notin B} \mu(B, y_1)\overline{P}(B|y_1),$$

which is non-negative as required. Thus $\overline{P}$ and $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ avoid sure loss.

For Problem 2, the analogous condition to (6) implies that $\pi(y_1|x_1) = 1$. Hence $\overline{P}(C|x_1) - I_C(y_1) = 1 - 1 = 0$ when $y_1 \in C$, so

$$\sum_{C \subseteq \mathcal{Y}} \tau(C, x_1)[\overline{P}(C|x_1) - I_C(y_1)] = \sum_{C \subseteq \mathcal{Y}, y_1 \notin C} \tau(C, x_1)\overline{P}(C|x_1) \geq 0.$$

Hence (5) holds for $x = x_1$ and $y = y_1$. Thus $\overline{P}$, $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ and $\{\overline{P}(\cdot|x)\colon x \in \mathcal{X}\}$ avoid sure loss. □

Given normal possibility measures $\overline{P}$ and $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$, let $\mathcal{M}_1$ denote the set of all probability measures (defined on the power set of $\mathcal{X} \times \mathcal{Y}$) which satisfy conditions (i) and (ii) in the proof of Theorem 1, i.e.

$$\mathcal{M}_1 = \{\, P \in \mathcal{P}\colon P(A) \leq \overline{P}(A) \text{ when } A \subseteq \mathcal{X} \times \mathcal{Y}, \text{ and } P(B \times \{y\}) \leq P(\mathcal{X} \times \{y\})\overline{P}(B|y) \text{ when } B \subseteq \mathcal{X}, y \in \mathcal{Y} \,\},$$

where $\mathcal{P}$ denotes the set of all probability measures whose domain is the power set of $\mathcal{X} \times \mathcal{Y}$. When $\mathcal{P}$ is regarded as a simplex in a finite-dimensional Euclidean space, its subset $\mathcal{M}_1$, which is the set of solutions for a finite system of weak linear inequalities, is a closed convex polyhedron. The proof of Theorem 1 shows that the given possibility measures avoid sure loss if and only if $\mathcal{M}_1$ is non-empty. The next two lemmas, which lead to Theorem 3, show that the given possibility measures are coherent if and only if they are 'upper envelopes' of $\mathcal{M}_1$, in a sense which is made explicit in Lemma 3.

Similarly, given normal possibility measures $\overline{P}$, $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ and $\{\overline{P}(\cdot|x)\colon x \in \mathcal{X}\}$, let $\mathcal{M}_2$ denote the set of all probability measures which satisfy conditions (i)–(iii) in the proof of Theorem 1. Again $\mathcal{M}_2$ is a closed convex polyhedron in $\mathcal{P}$, and the given possibility measures avoid sure loss if and only if $\mathcal{M}_2$ is non-empty.

Lemma 2. *Suppose that the possibility measures $\overline{P}$ and $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ avoid sure loss, so that $\mathcal{M}_1$ is non-empty.*

(α) *Condition* (7) *in the definition of coherence holds for a fixed set* $A_o \subseteq \mathcal{X} \times \mathcal{Y}$ *if and only if there is a probability measure* $P \in \mathcal{M}_1$ *such that* $P(A_o) = \overline{P}(A_o)$.

(β) *Condition* (8) *in the definition of coherence holds for a fixed* $B_o \subseteq \mathcal{X}$ *and* $v_o \in \mathcal{Y}$ *if and only if there is a probability measure* $P \in \mathcal{M}_1$ *such that* $P(B_o \times \{v_o\}) = P(\mathcal{X} \times \{v_o\})\overline{P}(B_o|v_o)$.

Proof. For (α), apply Lemma 1 to $\mathcal{D} = \{\lambda G(A)\colon \lambda \geq 0, A \subseteq \mathcal{X} \times \mathcal{Y}\} \cup \{\mu G(B|y)\colon \mu \geq 0, B \subseteq \mathcal{X}, y \in \mathcal{Y}\} \cup \{-G(A_o)\}$. Then condition (II) of Lemma 1 is equivalent to equation (7) for fixed A_o, and (I) of Lemma 1 is equivalent to the existence of a probability measure P which belongs to $\mathcal{M}_1$ and satisfies $P(G(A_o)) = 0$, i.e. $P(A_o) = \overline{P}(A_o)$. For ($\beta$), apply Lemma 1 to $\mathcal{D} = \{\lambda G(A)\colon \lambda \geq 0, A \subseteq \mathcal{X} \times \mathcal{Y}\} \cup \{\mu G(B|y)\colon \mu \geq 0, B \subseteq \mathcal{X}, y \in \mathcal{Y}\} \cup \{-G(B_o|v_o)\}$. □

Lemma 3. *Under the assumptions of Problems 1 and 2*

1. $\overline{P}$ *and* $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ *are coherent if and only if there is a non-empty set* $\mathcal{M}$ *of probability measures (each defined on the power set of* $\mathcal{X} \times \mathcal{Y}$*) such that*
   (i) $\overline{P}(A) = \sup\{P(A)\colon P \in \mathcal{M}\}$ *whenever* $A \subseteq \mathcal{X} \times \mathcal{Y}$, *and*
   (ii) $\overline{P}(B|y) \geq \sup\{P(B \times \{y\})/P(\mathcal{X} \times \{y\})\colon P \in \mathcal{M}, P(\mathcal{X} \times \{y\}) > 0\}$ *whenever* $B \subseteq \mathcal{X}$, $y \in \mathcal{Y}$ *and* $\overline{P}(\mathcal{X} \times \{y\}) > 0$, *with equality whenever* $\beta(y) < 1$, *where* $\beta(y) = \overline{P}(\mathcal{X} \times \mathrm{co}\{y\})$. *(There can be at most one value* $y \in \mathcal{Y}$ *for which* $\beta(y) < 1$.*)*

   If there is such a set $\mathcal{M}$ *then* $\mathcal{M}_1$ *is the largest such set.*
2. $\overline{P}$, $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ *and* $\{\overline{P}(\cdot|x)\colon x \in \mathcal{X}\}$ *are coherent if and only if there is a non-empty set* $\mathcal{M}$ *of probability measures which satisfies (i), (ii) and*
   (iii) $\overline{P}(C|x) \geq \sup\{P(\{x\} \times C)/P(\{x\} \times \mathcal{Y})\colon P \in \mathcal{M}, P(\{x\} \times \mathcal{Y}) > 0\}$ *whenever* $C \subseteq \mathcal{Y}$, $x \in \mathcal{X}$ *and* $\overline{P}(\{x\} \times \mathcal{Y}) > 0$, *with equality whenever* $\eta(x) < 1$, *where* $\eta(x) = \overline{P}(\mathrm{co}\{x\} \times \mathcal{Y})$.

   If there is such a set $\mathcal{M}$ *then* $\mathcal{M}_2$ *is the largest such set.*

Proof. We will prove part 1; the proof of part 2 is completely analogous. First suppose that a non-empty set $\mathcal{M}$ satisfies (i) and (ii). Then every $P \in \mathcal{M}$ must satisfy the inequalities in the definition of $\mathcal{M}_1$, so $\mathcal{M} \subseteq \mathcal{M}_1$. Hence $\overline{P}(A) = \sup\{P(A)\colon P \in \mathcal{M}\} \leq \sup\{P(A)\colon P \in \mathcal{M}_1\} \leq \overline{P}(A)$, which shows that $\mathcal{M}_1$ satisfies (i). Similarly $\mathcal{M}_1$ satisfies (ii). This proves that there is a set $\mathcal{M}$ which satisfies (i) and (ii) if and only if $\mathcal{M}_1$ does so. Since $\mathcal{M}$ is non-empty, so is $\mathcal{M}_1$, and the possibility measures avoid sure loss by Theorem 1. Also $\overline{P}(A_o) = \sup\{P(A_o)\colon P \in \mathcal{M}_1\}$ for any $A_o \subseteq \mathcal{X} \times \mathcal{Y}$, since $\mathcal{M}_1$ satisfies (i). Because $\mathcal{M}_1$ is a closed subset of $\mathcal{P}$ under the natural topology, the supremum value of $P(A_o)$ is attained on $\mathcal{M}_1$. Thus there is a probability measure $P \in \mathcal{M}_1$ such that $P(A_o) = \overline{P}(A_o)$. By ($\alpha$) of Lemma 2, the coherence condition (7) holds for every $A_o \subseteq \mathcal{X} \times \mathcal{Y}$. Also, for any $B_o \subseteq \mathcal{X}$ and $v_o \in \mathcal{Y}$, there is $P \in \mathcal{M}_1$ such that $P(B_o \times \{v_o\}) = P(\mathcal{X} \times \{v_o\})\overline{P}(B_o|v_o)$; this follows from (ii) when $\overline{P}(\mathcal{X} \times \mathrm{co}\{v_o\}) < 1$, since then $\overline{P}(B_o|v_o) = \sup\{P(B_o \times \{v_o\})/P(\mathcal{X} \times \{v_o\})\colon P \in \mathcal{M}_1\}$ and the supremum is attained by some $P \in \mathcal{M}_1$, and otherwise there is $P \in \mathcal{M}_1$ with $P(\mathcal{X} \times \mathrm{co}\{v_o\}) = \overline{P}(\mathcal{X} \times \mathrm{co}\{v_o\}) = 1$ and hence $P(\mathcal{X} \times \{v_o\}) = 0 = P(B_o \times \{v_o\})$. By ($\beta$) of Lemma 2, the coherence condition (8) holds for every $B_o \subseteq \mathcal{X}$ and $v_o \in \mathcal{Y}$. This establishes that $\overline{P}$ and $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ are coherent.

Conversely, suppose that $\overline{P}$ and $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ are coherent. Applying (α) of Lemma 2, for any $A \subseteq \mathcal{X} \times \mathcal{Y}$ there is $P \in \mathcal{M}_1$ such that $P(A) = \overline{P}(A)$. But $\sup\{P(A)\colon P \in \mathcal{M}_1\} \leq \overline{P}(A)$ by definition of $\mathcal{M}_1$, so there is equality and $\mathcal{M}_1$ satisfies (i). The inequality in (ii) is satisfied by $\mathcal{M}_1$, by definition of $\mathcal{M}_1$. To establish the equality (second part) of (ii), suppose that $B \subseteq \mathcal{X}$, $y \in \mathcal{Y}$ and $\overline{P}(\mathcal{X} \times \mathrm{co}\{y\}) < 1$. By ($\beta$) of Lemma 2, there is $P \in \mathcal{M}_1$ such that $P(B \times \{y\}) = P(\mathcal{X} \times \{y\})\overline{P}(B|y)$. Since $P \in \mathcal{M}_1$, $P(\mathcal{X} \times \mathrm{co}\{y\}) \leq \overline{P}(\mathcal{X} \times \mathrm{co}\{y\}) < 1$, so $P(\mathcal{X} \times \{y\}) = 1 - P(\mathcal{X} \times \mathrm{co}\{y\}) > 0$. Hence $P(B \times \{y\})/P(\mathcal{X} \times \{y\}) = \overline{P}(B|y)$, from which the second part of (ii) follows. Thus $\mathcal{M}_1$ satisfies (i) and (ii). □

Remark 7. Lemmas 2 and 3 remain valid when $\overline{P}$, $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ and $\{\overline{P}(\cdot|x)\colon x \in \mathcal{X}\}$ are general upper probability measures, e.g. possibility measures that are not normal. However, Theorem 3 relies on the assumption that $\overline{P}(\cdot|y)$ is a *coherent* upper probability measure (e.g. a *normal* possibility measure) when $\overline{P}(\mathcal{X} \times \mathrm{co}\{y\}) = 1$; this assumption is used in the following proof.

Proof of Theorem 3. Again we prove part 1 of the theorem; the proof of part 2 is analogous. First suppose that there are probability measures which satisfy (a)–(c) of Theorem 3. Let $\mathcal{M} = \{P_\gamma\colon \gamma \in \Gamma\}$, and verify that $\mathcal{M}$ satisfies conditions (i) and (ii) of Lemma 3. Clearly (a) implies (i). The first (inequality) part of (ii) holds because, whenever $P_\gamma(\mathcal{X}\times\{y\})>0$, $P_\gamma(B\times\{y\})/P_\gamma(\mathcal{X}\times\{y\}) = P_\gamma(B|y)\leq \overline{P}(B|y)$, using (c) and (b). Suppose that $\overline{P}(\mathcal{X}\times\mathrm{co}\{y\})<1$. Then, as in the proof of Lemma 3, $P_\gamma(\mathcal{X}\times\{y\})>0$ for all $\gamma\in\Gamma$. Applying (b) and (c), $\overline{P}(B|y)=\sup\{P_\gamma(B|y)\colon \gamma\in\Gamma\}=\sup\{P_\gamma(B\times\{y\})/P_\gamma(\mathcal{X}\times\{y\})\colon \gamma\in\Gamma\}$. This establishes the second (equality) part of (ii). Thus $\mathcal{M}$ satisfies (i) and (ii). By Lemma 3, $\overline{P}$ and $\{\overline{P}(\cdot|y)\colon y\in\mathcal{Y}\}$ are coherent. Thus (a)–(c) are sufficient for coherence.

Conversely, suppose that $\overline{P}$ and $\{\overline{P}(\cdot|y)\colon y\in\mathcal{Y}\}$ are coherent. By Lemma 3, the non-empty set $\mathcal{M}_1$ satisfies (i) and (ii). Let Γ index the probability measures in $\mathcal{M}_1$, i.e. $\mathcal{M}_1=\{P_\gamma\colon\gamma\in\Gamma\}$. (As explained below, a single probability measure $P\in\mathcal{M}_1$ may appear as P_γ with many different indices γ, to cope with the case where $P(\mathcal{X}\times\{y\})=0$.) For each $\gamma\in\Gamma$ and $y\in\mathcal{Y}$, define conditional probability measures $P_\gamma(\cdot|y)$ as follows. If $P_\gamma(\mathcal{X}\times\{y\})>0$ then define $P_\gamma(B|y)=P_\gamma(B\times\{y\})/P_\gamma(\mathcal{X}\times\{y\})$ whenever $B\subseteq\mathcal{X}$. This ensures that (c) is satisfied. Now consider the case $P_\gamma(\mathcal{X}\times\{y\})=0$. By assumption, $\overline{P}(\cdot|y)$ is a normal possibility measure and therefore a coherent upper probability measure. It follows that $\overline{P}(\cdot|y)$ can be written as the upper envelope of a finite set of probability measures $\{P^i_\gamma(\cdot|y)\colon i=1,2,\dots,k\}$, whose cardinality k is no greater than the cardinality of the power set of $\mathcal{X}$. Define the unconditional probability measure associated with $P^i_\gamma(\cdot|y)$ by $P^i_\gamma=P_\gamma$ for $i=1,2,\dots,k$. (That is, we include k copies of P_γ in the indexed set of measures, each copy being associated with a different conditional probability measure $P^i_\gamma(\cdot|y)$.)

As noted above, (c) holds by definition of $P_\gamma(\cdot|y)$. Finally, verify that conditions (a) and (b) are satisfied. By (i), $\overline{P}$ is the upper envelope of $\mathcal{M}_1=\{P_\gamma\colon\gamma\in\Gamma\}$, so (a) holds. If $\overline{P}(\mathcal{X}\times\mathrm{co}\{y\})<1$ then $P_\gamma(\mathcal{X}\times\{y\})>0$ for every $\gamma\in\Gamma$, so that $P_\gamma(B|y)=P_\gamma(B\times\{y\})/P_\gamma(\mathcal{X}\times\{y\})$, and (b) follows from (ii). If $\overline{P}(\mathcal{X}\times\mathrm{co}\{y\})=1$ then there is $P_\gamma\in\mathcal{M}_1$ with $P_\gamma(\mathcal{X}\times\{y\})=0$, by construction $\overline{P}(\cdot|y)$ is the upper envelope of $\{P_\gamma(\cdot|y)\colon P_\gamma(\mathcal{X}\times\{y\})=0,\gamma\in\Gamma\}$, and $\overline{P}(\cdot|y)$ dominates the upper envelope of $\{P_\gamma(\cdot|y)\colon P_\gamma(\mathcal{X}\times\{y\})>0,\gamma\in\Gamma\}$ by (c) and (ii). Hence (b) holds in all cases. Thus (a)–(c) are necessary for coherence. □

Lemma 4. *Suppose that π is a possibility distribution on a finite set Ω, and $\mathcal{M}$ is a set of probability measures (each defined on the power set of Ω) which satisfies, for all $\omega_o\in\Omega$, the two conditions:*

(a) $P(\{\omega\in\Omega\colon \pi(\omega)\leq\pi(\omega_o)\})\leq\pi(\omega_o)$ *for all* $P\in\mathcal{M}$*; and*
(b) *if* $\pi(\omega_o)>0$ *then there is* $P\in\mathcal{M}$ *such that* $P(\{\omega_o\})=\pi(\omega_o)$.

Then the possibility measure $\overline{P}$ generated by the possibility distribution π is the upper envelope of $\mathcal{M}$, i.e. $\overline{P}(A)=\max\{\pi(\omega)\colon\omega\in A\}=\max\{P(A)\colon P\in\mathcal{M}\}$ for all non-empty $A\subseteq\Omega$.

Proof. Suppose that $\mathcal{M}$ satisfies (a) and (b). Given a non-empty set $A\subseteq\Omega$, let ω_o maximise $\pi(\omega)$ over $\omega\in A$, so that $\pi(\omega_o)=\overline{P}(A)$. Then $A\subseteq\{\omega\in\Omega\colon\pi(\omega)\leq\pi(\omega_o)\}$. Using (a), we find that $P(A)\leq P(\{\omega\in\Omega\colon\pi(\omega)\leq\pi(\omega_o)\})\leq\pi(\omega_o)=\overline{P}(A)$ for all $P\in\mathcal{M}$. Thus $P(A)\leq\overline{P}(A)$ for all $P\in\mathcal{M}$.

If $\pi(\omega_o)=0$ then $0=\overline{P}(A)\geq P(A)\geq 0$, so $P(A)=0=\overline{P}(A)$ for all $P\in\mathcal{M}$. If $\pi(\omega_o)>0$ then, using (b), there is $P\in\mathcal{M}$ such that $P(\{\omega_o\})=\pi(\omega_o)=\overline{P}(A)$, hence $P(A)\geq P(\{\omega_o\})=\overline{P}(A)$. It follows that $\overline{P}(A)=\max\{P(A)\colon P\in\mathcal{M}\}$ in both cases. □

Proof of Theorem 4. First suppose that $\pi(\cdot,\cdot)$ and $\{\pi(\cdot|y)\colon y\in\mathcal{Y}\}$ are coherent. The lower and upper bounds for $\pi(x|y)$ can be obtained directly from the two inequalities in the definition of coherence in Section 4.2. Let $u_o\in\mathcal{X}$ and $v_o\in\mathcal{Y}$. For the lower bound, substitute into (7) the values $A_o=\{(u_o,v_o)\}$, $\lambda(\mathcal{X}\times\{v_o\})=\pi(u_o|v_o)$, $\lambda(A)=0$ for all other $A\subseteq\mathcal{X}\times\mathcal{Y}$, $\mu(\{u_o\},v_o)=1$ and $\mu(B,v)=0$ for all other $B\subseteq\mathcal{X}$ and $v\in\mathcal{Y}$. Using $\overline{P}(\mathcal{X}\times\{v_o\})=\pi(v_o)$, $\overline{P}(\{u_o\}|v_o)=\pi(u_o|v_o)$ and $\overline{P}(\{(u_o,v_o)\})=\pi(u_o,v_o)$ and cancelling the terms involving indicator functions, (7) implies that $\pi(u_o|v_o)\pi(v_o)\geq\pi(u_o,v_o)$. This gives the lower bound $\pi(u_o,v_o)/\pi(v_o)$ for $\pi(u_o|v_o)$, if $\pi(v_o)>0$.

To obtain the upper bound, substitute into (8) $B_o=\{u_o\}$, $\lambda(\{(u_o,v_o)\})=1-\pi(u_o|v_o)$, $\lambda(\{(u_o,v_o)\}\cup\mathcal{X}\times\mathrm{co}\{v_o\})=\pi(u_o|v_o)$, $\lambda(A)=0$ for all other $A\subseteq\mathcal{X}\times\mathcal{Y}$, and $\mu(B,v)=0$ for all $B\subseteq\mathcal{X}$ and $v\in\mathcal{Y}$. Using $\overline{P}(\{(u_o,v_o)\})=\pi(u_o,v_o)$, $\overline{P}(\{(u_o,v_o)\}\cup\mathcal{X}\times\mathrm{co}\{v_o\})=\max\{\pi(u_o,v_o),\beta(v_o)\}$ where $\beta(v_o)=\overline{P}(\mathcal{X}\times\mathrm{co}\{v_o\})=\max\{\pi(v)\colon v\in\mathcal{Y},v\neq v_o\}$, and $\overline{P}(\{u_o\}|v_o)=\pi(u_o|v_o)$, and cancelling the two terms involving $I_{\{(u_o,v_o)\}}$, (8) implies that (for some x, y)

$$[1-\pi(u_o|v_o)]\pi(u_o,v_o)+\pi(u_o|v_o)[\max\{\pi(u_o,v_o),\beta(v_o)\}-I_{\mathcal{X}\times\mathrm{co}\{v_o\}}]-\pi(u_o|v_o)I_{\mathcal{X}\times\{v_o\}}\geq 0.$$

Combining the two indicator functions and rearranging terms, $\pi(u_o,v_o)\geq\pi(u_o|v_o)[\pi(u_o,v_o)+1-\max\{\pi(u_o,v_o),\beta(v_o)\}]$. If the term in square brackets is zero then $\pi(u_o,v_o)=0$ and $\beta(v_o)=1$, in which case the upper bound in the theorem is the trivial value 1. Otherwise we can divide both sides by the term in square brackets to obtain the upper bound for $\pi(u_o|v_o)$.

Conversely, suppose that $\pi(\cdot,\cdot)$ and $\{\pi(\cdot|y)\colon y\in\mathcal{Y}\}$ satisfy the conditions of Theorem 4. To prove that they are coherent, we will construct a set $\mathcal{M}$ of probability measures which satisfy the two conditions of Lemma 3. For each $y\in\mathcal{Y}$, let $x=u(y)$ maximise $\pi(x,y)$, so that $\pi(u(y),y)=\pi(y)$. Let $y_1\in\mathcal{Y}$ be such that $\pi(y_1)=1$, and let $y_2\in\mathrm{co}\{y_1\}$ maximise $\pi(y)$ over all $y\in\mathrm{co}\{y_1\}$.

Suppose that $\pi(x,y)>0$. Then $\pi(y)>0$, and $\pi(x|y)>0$ by the lower bound in Theorem 4. Define a probability measure $P_{x,y}$ to have the following probability mass function: $P_{x,y}(x,y)=\pi(x,y)$, $P_{x,y}(u(y),y)=\pi(x,y)[1-\pi(x|y)]/\pi(x|y)$ if $x\neq u(y)$, $P_{x,y}(u(y_1),y_1)=1-\pi(x,y)/\pi(x|y)$ if $y\neq y_1$, $P_{x,y}(u(y_2),y_2)=1-\pi(x,y_1)/\pi(x|y_1)$ if $y=y_1$ and $P_{x,y}(u,v)=0$ for all other $(u,v)\in\mathcal{X}\times\mathcal{Y}$. Then $P_{x,y}(u,v)>0$ for at most three pairs (u,v), i.e. for $(u,v)=(x,y)$, $(u(y),y)$ and either $(u(y_1),y_1)$ or $(u(y_2),y_2)$. It can be easily verified that these probabilities are non-negative, using the lower bound in the theorem to show that $\pi(x,y)\leq\pi(x|y)$, and that they sum to one, using the lower bound again to show that $\pi(u(y)|y)=1$. Thus each $P_{x,y}$ is a probability measure. Define $\mathcal{M}=\{P_{x,y}\colon(x,y)\in\mathcal{X}\times\mathcal{Y},\pi(x,y)>0\}$.

Next verify that $\mathcal{M}$ satisfies condition (i) of Lemma 3, i.e. that $\mathcal{M}$ has upper envelope $\overline{P}$ where $\overline{P}(A)=\max\{\pi(u,v)\colon(u,v)\in A\}$ for $A\subseteq\mathcal{X}\times\mathcal{Y}$. By Lemma 4, it suffices to prove that

(a) $P_{x,y}(\{(u,v)\in\mathcal{X}\times\mathcal{Y}\colon\pi(u,v)\leq\pi(w,z)\}\leq\pi(w,z)$ whenever $(w,z)\in\mathcal{X}\times\mathcal{Y}$ and $P_{x,y}\in\mathcal{M}$; and
(b) if $(x,y)\in\mathcal{X}\times\mathcal{Y}$ and $\pi(x,y)>0$ then there is $P\in\mathcal{M}$ such that $P(x,y)=\pi(x,y)$.

Clearly (b) is satisfied, since $P_{x,y}(x,y)=\pi(x,y)$ whenever $(x,y)\in\mathcal{X}\times\mathcal{Y}$ and $\pi(x,y)>0$, by definition of $P_{x,y}$.

Consider (a). Suppose that $(x,y)\in\mathcal{X}\times\mathcal{Y}$ and $\pi(x,y)>0$. It suffices to verify (a) for those (w,z) which have positive probability under $P_{x,y}$, and therefore we need only to consider $(w,z)\in\{(x,y),(u(y),y),(u(y_1),y_1),(u(y_2),y_2)\}$. Clearly (a) holds when $(w,z)=(u(y_1),y_1)$, since $\pi(u(y_1),y_1)=1$. When $y\neq y_1$, we need to consider only $(w,z)=(x,y)$ and $(w,z)=(u(y),y)$: (a) holds in the first case because $P_{x,y}(x,y)=\pi(x,y)$, and also in the second case because, using the lower bound in the theorem, $P_{x,y}(\{(x,y),(u(y),y)\})=\pi(x,y)/\pi(x|y)\leq\pi(y)=\pi(u(y),y)$.

Suppose that $y=y_1$. We need to consider only $(w,z)=(x,y_1)$ and $(w,z)=(u(y_2),y_2)$. Using the upper bound for $\pi(x|y_1)$ in the theorem, $\pi(x,y_1)/\pi(x|y_1)\geq\pi(x,y_1)+1-\max\{\pi(x,y_1),\beta(y_1)\}$, where $\beta(y_1)=\pi(y_2)=\pi(u(y_2),y_2)$ since y_2 maximises $\pi(y)$ over $y\neq y_1$. (This holds when $\beta(y_1)=1$ as then $\pi(x|y_1)\leq 1$.) Consequently, $P_{x,y}(\{(x,y_1),(u(y_2),y_2)\})=\pi(x,y_1)+1-\pi(x,y_1)/\pi(x|y_1)\leq\max\{\pi(x,y_1),\pi(u(y_2),y_2)\}$. Using $\max\{\pi(x,y_1),\beta(y_1)\}\leq\pi(x,y_1)+\beta(y_1)=\pi(x,y_1)+\pi(y_2)$, the first inequality gives $\pi(x,y_1)/\pi(x|y_1)\geq 1-\pi(y_2)$, from which it follows that $P_{x,y}(u(y_2),y_2)=1-\pi(x,y_1)/\pi(x|y_1)\leq\pi(y_2)$. Finally, $P_{x,y}(x,y)=\pi(x,y)$. This establishes (a). Thus $\mathcal{M}$ satisfies (i) of Lemma 3.

Next verify (ii) of Lemma 3. Suppose that $\pi(x,y)>0$ and $v\in\mathcal{Y}$. If $P_{x,y}(\mathcal{X}\times\{v\})>0$, define a conditional probability measure $P_{x,y}(\cdot|v)$ by $P_{x,y}(B|v)=P_{x,y}(B\times\{v\})/P_{x,y}(\mathcal{X}\times\{v\})$ for all $B\subseteq\mathcal{X}$. We need to prove that $P_{x,y}(B|v)\leq\overline{P}(B|v)$ whenever the left-hand side is defined.

First consider the case $v=y$. Then $P_{x,y}(\cdot|y)$ is defined since $P_{x,y}(\mathcal{X}\times\{y\})\geq\pi(x,y)>0$. To verify that $P_{x,y}(B|y)\leq\overline{P}(B|y)$, we need only to consider sets B that contain x or $u(y)$, since $P_{x,y}(u,y)=0$ for all other values of u. The inequality holds when $u(y)\in B$ because, using the lower bound in the theorem, $\overline{P}(B|y)\geq\pi(u(y)|y)\geq\pi(u(y),y)/\pi(y)=1$. It holds also when B contains x but not $u(y)$, because then $P_{x,y}(B|y)=P_{x,y}(\{x\}|y)=P_{x,y}(x,y)/[P_{x,y}(x,y)+P_{x,y}(u(y),y)]=\pi(x|y)=\overline{P}(\{x\}|y)\leq\overline{P}(B|y)$ (using the definition of $P_{x,y}$). (This holds when $x=u(y)$, using the lower bound in the theorem, as then all terms are 1.) Thus the inequality holds for $v=y$.

The remaining cases for which $P_{x,y}(\cdot|v)$ may be defined are the cases $v=y_1\neq y$ and $v=y_2\neq y$. The inequality holds trivially in these cases because, using the lower bound in the theorem, $\overline{P}(B|y_i)=$

$\pi(u(y_i)|y_i) = 1$ if $u(y_i) \in B$, and otherwise $P_{x,y}(B|y_i) = 0$. This establishes that $P_{x,y}(B|v) \leq \overline{P}(B|v)$ whenever the left-hand side is defined, whence $\overline{P}(B|v) \geq \sup\{P_{x,y}(B|v)\colon P_{x,y} \in \mathcal{M}, P_{x,y}(\mathcal{X} \times \{v\}) > 0\}$ whenever $B \subseteq \mathcal{X}$, $v \in \mathcal{Y}$ and $\overline{P}(\mathcal{X} \times \{v\}) > 0$. To complete the proof of (ii), we need to show that there is equality whenever $\beta(v) < 1$, where $\beta(v) = \overline{P}(\mathcal{X} \times \mathrm{co}\{v\})$.

Suppose that $\beta(v) < 1$. (That can happen only when $v = y_1$.) Given $B \subseteq \mathcal{X}$, let $u = x$ maximise $\pi(u|v)$ over $u \in B$, so that $\overline{P}(B|v) = \max\{\pi(u|v)\colon u \in B\} = \pi(x|v)$. Now $P(\cdot|v)$ is defined for all $P \in \mathcal{M}$ since $P(\mathcal{X} \times \{v\}) \geq 1 - \beta(v) > 0$. To establish the required equality, it suffices to prove that $P(\{x\}|v) \geq \pi(x|v)$ for some $P \in \mathcal{M}$, because then $P(B|v) \geq P(\{x\}|v) \geq \pi(x|v) = \overline{P}(B|v)$. This holds trivially when $\pi(x, v) = 0$, since then $\pi(x|v) = 0$ by the upper bound in the theorem, using $\beta(v) < 1$. It holds also when $\pi(x, v) > 0$, because then $P_{x,v}$ is defined and $P_{x,v}(\{x\}|v) = \pi(x|v)$. This completes the proof of (ii).

By Lemma 3, $\overline{P}$ and $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ are coherent, i.e. the corresponding possibility distributions $\pi(\cdot, \cdot)$ and $\{\pi(\cdot|y)\colon y \in \mathcal{Y}\}$ are coherent. □

Proof of Theorem 5. Let $\overline{P}$ be the upper probability measure that is generated by $\pi(\cdot, \cdot)$, and let $\mathcal{M}$ denote the set of all probability measures that are dominated by $\overline{P}$, i.e.

$$\mathcal{M} = \{P \in \mathcal{P}\colon P(A) \leq \overline{P}(A) \text{ for all } A \subseteq \mathcal{X} \times \mathcal{Y}\}.$$

Verify that $\mathcal{M}$ satisfies the conditions of Lemma 3. Condition (i) holds because $\overline{P}$ is a coherent upper probability measure, which implies that $\overline{P}$ is the upper envelope of $\mathcal{M}$ by Theorem 3.3.3 of [24]. Condition (ii) holds trivially whenever $\beta(y) = 1$, since then $\pi_{NE}(\cdot|y)$ is vacuous and $\overline{P}(B|y) = \max\{\pi_{NE}(x|y)\colon x \in B\} = 1$ whenever B is non-empty.

Suppose that $\beta(y) < 1$. Then, for every $P \in \mathcal{M}$, $P(\mathcal{X} \times \{y\}) > 0$ and $P(\cdot|y)$ is defined by Bayes' rule. To establish (ii) of Lemma 3 we need to prove that, for every non-empty $B \subseteq \mathcal{X}$,

(a) $P(B|y) \leq \overline{P}(B|y)$ whenever $P \in \mathcal{M}$, and
(b) for some $P \in \mathcal{M}$, $P(B|y) = \overline{P}(B|y)$.

Let $u = x$ maximise $\pi(u, y)$ over $u \in B$, so that $\overline{P}(B \times \{y\}) = \pi(x, y)$ and moreover $\overline{P}(B|y) = \max\{\pi_{NE}(u|y)\colon u \in B\} = \pi_{NE}(x|y)$, using the fact that $\pi_{NE}(u|y)$ is non-decreasing in $\pi(u, y)$. To prove (a), suppose that $P \in \mathcal{M}$. Then $P(B \times \{y\}) \leq \overline{P}(B \times \{y\}) = \pi(x, y)$, and $P(\mathrm{co}B \times \{y\}) = 1 - P(B \times \{y\} \cup \mathcal{X} \times \mathrm{co}\{y\}) \geq 1 - \overline{P}(B \times \{y\} \cup \mathcal{X} \times \mathrm{co}\{y\}) = 1 - \max\{\overline{P}(B \times \{y\}), \overline{P}(\mathcal{X} \times \mathrm{co}\{y\})\} = 1 - \max\{\pi(x, y), \beta(y)\}$. Hence $P(B|y) = P(B \times \{y\})/[P(B \times \{y\}) + P(\mathrm{co}B \times \{y\})] \leq \pi(x, y)/[\pi(x, y) + 1 - \max\{\pi(x, y), \beta(y)\}] = \pi_{NE}(x|y) = \overline{P}(B|y)$.

To prove (b), let $y_2 \in \mathcal{Y}$ satisfy $\pi(y_2) = \beta(y)$ and $y_2 \neq y$, and let $u(v) \in \mathcal{X}$ satisfy $\pi(u(v), v) = \pi(v)$. Define a probability measure P to have probability mass function $P(x, y) = \pi(x, y)$, $P(u(y), y) = 1 - \max\{\pi(x, y), \beta(y)\}$, $P(u(y_2), y_2) = \max\{0, \beta(y) - \pi(x, y)\}$, and $P(w, z) = 0$ for all other $(w, z) \in \mathcal{X} \times \mathcal{Y}$. (This agrees with the probability measure $P_{x,y}$ defined in the proof of Theorem 4, provided that $\pi(x, y) > 0$.) Then clearly $P(B|y) \geq P(\{x\}|y) = \pi(x, y)/[\pi(x, y) + 1 - \max\{\pi(x, y), \beta(y)\}] = \pi_{NE}(x|y) = \overline{P}(B|y)$.

It remains to be shown that $P \in \mathcal{M}$, i.e. that $P(A) \leq \overline{P}(A)$ for all $A \subseteq \mathcal{X} \times \mathcal{Y}$, and it suffices to verify this inequality when A is a subset of $\{(x, y), (u(y), y), (u(y_2), y_2)\}$ because only these elements can have positive probability under P. The inequality holds when $(u(y), y) \in A$ because then $\overline{P}(A) \geq \pi(u(y), y) = \pi(y) = 1$, since $\beta(y) < 1$. It holds for $A = \{(x, y)\}$ since $P(x, y) = \pi(x, y)$, and for $A = \{(u(y_2), y_2)\}$ since $P(u(y_2), y_2) \leq \beta(y) = \pi(u(y_2), y_2)$. Finally, the inequality holds when $A = \{(x, y), (u(y_2), y_2)\}$ because then $P(A) = \max\{\pi(x, y), \beta(y)\} = \overline{P}(A)$. This completes the verification of (ii) in Lemma 3.

Condition (iii) of Lemma 3 holds similarly, since both $\pi_{NE}(\cdot|\cdot)$ and $\mathcal{M}$ are defined symmetrically in x and y (unlike the $\mathcal{M}$ used in the proof of Theorem 4). It follows from part 2 of Lemma 3 that $\overline{P}$, $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ and $\{\overline{P}(\cdot|x)\colon x \in \mathcal{X}\}$ are coherent, i.e. the corresponding possibility distributions $\pi(\cdot, \cdot)$, $\{\pi_{NE}(\cdot|y)\colon y \in \mathcal{Y}\}$ and $\{\pi_{NE}(\cdot|x)\colon x \in \mathcal{X}\}$ are coherent. □

Remark 8. As in Theorem 4, if $\pi(y) = 0$ or $\pi(x) = 0$ then $\pi(\cdot|y)$ or $\pi(\cdot|x)$ can be taken to be any possibility distribution without affecting coherence. A slight modification of the preceding proof establishes that coherence holds also for regular extension in Problem 2. The only part of the proof that needs to be modified is the verification of $P(B|y) \leq \overline{P}(B|y)$ when $\beta(y) = 1$, $\overline{P}(B|y) = 0$ and $P(\mathcal{X} \times \{y\}) > 0$. Then

$\pi_{RE}(x|y) = 0$ for all $x \in B$. This implies that $0 = \pi(x,y) \geq P(x,y)$ for all $x \in B$, and hence that $P(B \times \{y\}) = 0$ and $P(B|y) = 0$, which establishes the required inequality.

Proof of Corollary to Theorem 5. From the proof of Theorem 5 we deduce that, when $\beta(y) < 1$, $\overline{P}(B|y) = \max\{P(B|y)\colon P \in \mathcal{M}\}$, where $P(B|y) = P(B \times \{y\})/P(\mathcal{X} \times \{y\})$ and $\mathcal{M}$ is the set of all probability measures that are dominated by $\overline{P}$. The result therefore holds when $\beta(y) < 1$.

Suppose that $\beta(y) = 1$. Then there are $(u,v) \in \mathcal{X} \times \mathcal{Y}$ such that $v \neq y$ and $\pi(u,v) = 1$. Consider the degenerate probability measure P such that $P(u,v) = 1$. Clearly P belongs to $\mathcal{M}$. But $P(\mathcal{X} \times \{y\}) = 0$, so P is coherent with *any* conditional probability measure $P(\cdot|y)$. The upper envelope of the set of all conditional probability measures is the vacuous upper probability $\overline{P}(B|y) = 1$ for all non-empty $B \subseteq \mathcal{X}$, which corresponds to the vacuous conditional possibility distribution $\pi_{NE}(x|y) = 1$ for all $x \in \mathcal{X}$. □

Proof of Theorem 6. Let $\overline{P}$ be the possibility measure generated by $\pi(\cdot,\cdot)$. By the second part of Lemma 3, it suffices to construct a set of probability measures, $\mathcal{M}$, which has upper envelope $\overline{P}$ and also satisfies conditions (ii) and (iii) of Lemma 3, which relate $\mathcal{M}$ to the conditional possibility measures generated by $\pi_{HM}(\cdot|\cdot)$. There are four steps in the proof:

(a) Given $(x,y) \in \mathcal{X} \times \mathcal{Y}$ such that $\pi(x,y) > 0$, construct a probability measure $P_{x,y}$ with $P_{x,y}(x,y) = \pi(x,y)$. We then define $\mathcal{M} = \{P_{x,y}\colon x \in \mathcal{X}, y \in \mathcal{Y}, \pi(x,y) > 0\}$.

(b) Verify that each $P_{x,y}$ is dominated by the possibility measure $\overline{P}$. By Lemma 4, it is enough to show that $P_{x,y}(\{(u,v)\colon \pi(u,v) \leq \pi(w,z)\}) \leq \pi(w,z)$ whenever $(w,z) \in \mathcal{X} \times \mathcal{Y}$. It then follows from Lemma 4 that $\overline{P}$ is the upper envelope of $\mathcal{M}$, i.e. condition (i) of Lemma 3 holds.

(c) Verify that $P_{x,y}(B|v) \leq \overline{P}(B|v) = \max\{\pi_{HM}(u|v)\colon u \in B\}$ whenever $B \subseteq \mathcal{X}$, $v \in \mathcal{Y}$, $P_{x,y}(\mathcal{X} \times \{v\}) > 0$ and the conditional probabilities are defined by Bayes' rule. Applying Lemma 4(a) to the conditional probability measure $P_{x,y}(\cdot|v)$ and the conditional possibility distribution $\pi_{HM}(\cdot|v)$, it is enough to show that $P_{x,y}(\{u\colon u \in \mathcal{X}, \pi_{HM}(u|v) \leq \pi_{HM}(w|v)\}|v) \leq \pi_{HM}(w|v)$ whenever $P_{x,y}(\mathcal{X} \times \{v\}) > 0$ and $w \in \mathcal{X}$. Similarly verify that $P_{x,y}(C|u) \leq \overline{P}(C|u)$ whenever $C \subseteq \mathcal{Y}$, $u \in \mathcal{X}$ and $P_{x,y}(\{u\} \times \mathcal{Y}) > 0$.

(d) Show that $P_{x,y}(x|y) = \pi_{HM}(x|y)$ whenever $\pi_{HM}(x|y) > 0$ and $\beta(y) < 1$ (which imply $\pi(x,y) > 0$), and similarly $P_{x,y}(y|x) = \pi_{HM}(y|x)$ whenever $\pi_{HM}(y|x) > 0$ and $\eta(x) < 1$. (In fact we can define $P_{x,y}$ so that $P_{x,y}(x|y) = \pi_{HM}(x|y)$ and $P_{x,y}(y|x) = \pi_{HM}(y|x)$ whenever $\pi(x,y) > 0$.) Using Lemma 4, it follows from (c) and (d) that conditions (ii) and (iii) of Lemma 3 hold.

It then follows from Lemma 3 that $\overline{P}$, $\{\overline{P}(\cdot|y)\colon y \in \mathcal{Y}\}$ and $\{\overline{P}(\cdot|x)\colon x \in \mathcal{X}\}$ are coherent, i.e. the corresponding possibility distributions $\pi(\cdot,\cdot)$, $\{\pi_{HM}(\cdot|y)\colon y \in \mathcal{Y}\}$ and $\{\pi_{HM}(\cdot|x)\colon x \in \mathcal{X}\}$ are coherent.

We assume throughout that $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $\pi(x,y) > 0$. The first step is to construct $P_{x,y}$ to satisfy $P_{x,y}(x,y) = \pi(x,y)$. In the proof of Theorem 4, that was done simply by assigning positive probability to (u_2,y) where $u = u_2$ maximises $\pi(u,y)$, and assigning all the remaining probability to a single element of $\mathcal{X} \times \mathcal{Y}$. The positive probability of (u_2,y) will often produce $P(y|u_2) = 1$. That caused no difficulty in the proof of Theorem 4 because in Problem 1 we were not concerned with $\pi(y|x)$. However Problem 2 is more complicated, and in particular we need to ensure that $P_{x,y}(y|u_2) \leq \pi_{HM}(y|u_2)$. To do so, we may need to assign positive probability to (u_2,v_3) where $v = v_3$ maximises $\pi(u_2,v)$. But then, to ensure that $P_{x,y}(u_2|v_3) \leq \pi_{HM}(u_2|v_3)$, we may need to assign positive probability to some (u_4,v_3), and so on, producing a sequence of terms (x,y), (u_2,y), (u_2,v_3), (u_4,v_3), (u_4,v_5), … with positive probability. Similarly, to ensure that $P_{x,y}(y|x) \leq \pi_{HM}(y|x)$, we typically need to assign positive probability to (x,v_2) for some $v_2 \neq y$, and hence to all terms in a second sequence (x,y), (x,v_2), (u_3,v_2), (u_3,v_4), …

Formally, write $u_o = u_1 = x$ and $v_o = v_1 = y$, let u_i maximise $\pi(u,v_{i-1})$ for $i = 2,3,\ldots$ and let v_j maximise $\pi(u_{j-1},v)$ for $j = 2,3,\ldots$, where each sequence terminates when no further increase in $\pi(u,v)$ is possible. (If there are several maximising values of u or v, choose the one which appears first in a fixed ordering of $\mathcal{X}$ or $\mathcal{Y}$.) Both sequences must terminate because $\mathcal{X} \times \mathcal{Y}$ is finite and $\pi(u,v)$ is strictly increasing along each sequence. Denoting the termination points (which may or may not be different) by

(s_1,t_1) and (s_2,t_2),

$$\pi(x,y)=\pi(u_o,v_1)<\pi(v_1)=\pi(u_2,v_1)<\pi(u_2)=\pi(u_2,v_3) \\ <\pi(v_3)=\pi(u_4,v_3)<\cdots<\pi(s_1)=\pi(t_1)=\pi(s_1,t_1)$$

and

$$\pi(x,y)=\pi(u_1,v_o)<\pi(u_1)=\pi(u_1,v_2)<\pi(v_2)=\pi(u_3,v_2) \\ <\pi(u_3)=\pi(u_3,v_4)<\cdots<\pi(s_2)=\pi(t_2)=\pi(s_2,t_2).$$

The probability measure $P_{x,y}$ is defined to give the following probability masses to the terms in the two sequences:

$$P_{x,y}(x,y)=\pi(x,y),$$
$$P_{x,y}(u_i,v_{i+1})=\frac{1}{2}[\pi(u_i,v_{i+1})-\pi(u_i,v_{i-1})],$$
$$P_{x,y}(u_{i+1},v_i)=\frac{1}{2}[\pi(u_{i+1},v_i)-\pi(u_{i-1},v_i)]$$

for $i=1,2,\dots$

Although the terms in one sequence must be distinct, since $\pi(u,v)$ is strictly increasing, one term $(u,v)\neq(x,y)$ may appear in both sequences. If so, take its probability to be the sum of the two probabilities defined in the previous paragraph. Then the total probability assigned to the two sequences is $\pi(x,y)+\frac{1}{2}[\pi(u_2,v_1)-\pi(x,y)+\pi(u_2,v_3)-\pi(u_2,v_1)+\cdots+\pi(s_1,t_1)+\pi(u_1,v_2)-\pi(x,y)+\pi(u_3,v_2)-\pi(u_1,v_2)+\cdots+\pi(s_2,t_2)]=\frac{1}{2}[\pi(s_1,t_1)+\pi(s_2,t_2)]\leq 1$. (The factor $\frac{1}{2}$ enables us to handle both sequences simultaneously.) Any remaining probability $1-\frac{1}{2}[\pi(s_1,t_1)+\pi(s_2,t_2)]$ needs to be assigned to other elements of $\mathcal{X}\times\mathcal{Y}$.

There are five cases to consider, depending on the values of $\pi(x,y)$, $\pi(x)$, $\pi(y)$, $\eta(x)$ and $\beta(y)$. In each case we complete the definition of $P_{x,y}$ and outline the four steps of the proof, omitting some details.

Case (i). $\pi(x)<1$ and $\pi(y)<1$. (This implies that $\pi(x,y)<\beta(y)=\eta(x)=1$.)

(a) Since in this case $\pi(x)=\max\{\pi(x,v)\colon v\in\mathcal{Y}\}<1$, $\pi(y)=\max\{\pi(u,y)\colon u\in\mathcal{X}\}<1$ and in general $\max\{\pi(u,v)\colon (u,v)\in\mathcal{X}\times\mathcal{Y}\}=1$, there is $(q,r)\in\mathcal{X}\times\mathcal{Y}$ such that $q\neq x$, $r\neq y$ and $\pi(q,r)=1$. Assign all the remaining probability $1-\frac{1}{2}[\pi(s_1,t_1)+\pi(s_2,t_2)]$ to (q,r), and let $P_{x,y}(u,v)=0$ if (u,v) is not in either sequence and $(u,v)\neq(q,r)$. (If possible, take (q,r) to be the termination point of either sequence.)

(b) Let $(w,z)\in\mathcal{X}\times\mathcal{Y}$ and $A(w,z)=\{(u,v)\colon \pi(u,v)\leq\pi(w,z)\}$. We need to show that $P_{x,y}(A(w,z))\leq\pi(w,z)$. This obviously holds if $\pi(w,z)=1$. If $\pi(u,v)\leq\pi(x,y)$ and $(u,v)\neq(x,y)$ then $P_{x,y}(u,v)=0$ because (u,v) cannot belong to either sequence. Hence, if $\pi(w,z)\leq\pi(x,y)$ and $(w,z)\neq(x,y)$ then $P_{x,y}(A(w,z))=0\leq\pi(w,z)$. If $(w,z)=(x,y)$ then $P_{x,y}(A(w,z))=\pi(w,z)$.

Finally, if $\pi(x,y)\leq\pi(w,z)<1$ then $P_{x,y}(A(w,z))=\pi(x,y)+\sum P_{x,y}(u,v)$, where the summation is over all terms (u,v) in the two sequences for which $\pi(x,y)<\pi(u,v)\leq\pi(z,w)$. Let (w_1,z_1) and (w_2,z_2) denote the last terms in each sequence for which $\pi(w_i,z_i)\leq\pi(w,z)$. Then $P_{x,y}(A(w,z))=\pi(x,y)+\frac{1}{2}[\pi(w_1,z_1)-\pi(x,y)]+\frac{1}{2}[\pi(w_2,z_2)-\pi(x,y)]=\frac{1}{2}[\pi(w_1,z_1)+\pi(w_2,z_2)]\leq\pi(w,z)$.

(c) Consider conditioning on $v\in\mathcal{Y}$ such that $P_{x,y}(\mathcal{X}\times\{v\})>0$. (Conditioning on $u\in\mathcal{X}$ is analogous because case (i) is symmetric between $\mathcal{X}$ and $\mathcal{Y}$.) We need to verify that, if we define the conditional probabilities by Bayes' rule, then $P_{x,y}(B|v)\leq\overline{P}(B|v)$ whenever $B\subseteq\mathcal{X}$. From the construction of $P_{x,y}$, there are at most three elements u in $\mathcal{X}$ such that $P_{x,y}(u,v)>0$, and one of these (denoted by w) must achieve $\pi(w,v)=\pi(v)$. The result is trivial when $w\in B$, since then $\overline{P}(B|v)=\pi_{HM}(w|v)=1$.

Hence we need to verify the result only when B contains one or two elements u such that $\pi(u,v)<\pi(w,v)=\pi(v)$ and $P_{x,y}(u,v)>0$. In that case (u,v) must appear in one of the sequences and must be followed by (w,v), so that $P_{x,y}(w,v)\geq\frac{1}{2}[\pi(w,v)-\pi(u,v)]=\frac{1}{2}[\pi(v)-\pi(u,v)]$. If $(u,v)\neq(x,y)$ then $P_{x,y}(u,v)\leq\frac{1}{2}\pi(u,v)$, hence (using Bayes' rule) $P_{x,y}(u|v)=P_{x,y}(u,v)/P_{x,y}(v)\leq P_{x,y}(u,v)/[P_{x,y}(u,v)+P_{x,y}(w,v)]\leq\pi(u,v)/[\pi(u,v)+\pi(v)-\pi(u,v)]=\pi(u,v)/\pi(v)\leq\pi_{HM}(u|v)=$

$\overline{P}(\{u\}|v)$. If $(u,v) = (x,y)$ then $P_{x,y}(x,y) = \pi(x,y)$, hence $P_{x,y}(x|y) = P_{x,y}(x,y)/P_{x,y}(y) \leq P_{x,y}(x,y)/[P_{x,y}(x,y) + P_{x,y}(w,y)] \leq 2\pi(x,y)/[\pi(y) + \pi(x,y)] = \pi_{HM}(x|y) = \overline{P}(\{x\}|y)$, using the fact that $\beta(y) = 1$. This establishes the result when B contains only one element.

Finally consider $B = \{u,u'\}$, where both (u,v) and (u',v) have positive probability under $P_{x,y}$, $\pi(u,v) < \pi(w,v) = \pi(v)$ and similarly for u'. Then (u,v) and (u',v) must appear in different sequences, and each must be followed by (w,v). Hence $P_{x,y}(w,v) \geq \frac{1}{2}[\pi(w,v)-\pi(u,v)]+\frac{1}{2}[\pi(w,v)-\pi(u',v)] = \pi(v) - \frac{1}{2}[\pi(u,v)+\pi(u',v)]$. Also $P_{x,y}(\{(u,v),(u',v)\}) \leq \frac{1}{2}[\pi(u,v)+\pi(u',v)]$; this is true when $(u',v) = (x,y)$ because then $P_{x,y}(u,v) \leq \frac{1}{2}[\pi(u,v) - \pi(x,y)]$. As a result, we have that $P_{x,y}(\{u,u'\}|v) = P_{x,y}(\{(u,v),(u',v)\})/P_{x,y}(\{(u,v),(u',v),(w,v)\}) \leq \frac{1}{2}[\pi(u,v) + \pi(u',v)]/\pi(v) \leq \max\{\pi(u,v)/\pi(v), \pi(u',v)/\pi(v)\} \leq \max\{\pi_{HM}(u|v), \pi_{HM}(u'|v)\} = \overline{P}(\{u,u'\}|v)$.

(d) This holds trivially in case (i) because $\beta(y) = \eta(x) = 1$. However, the definition of $P_{x,y}$ in case (i) can easily be modified to ensure that $P_{x,y}(x|y) = \pi_{HM}(x|y)$ and $P_{x,y}(y|x) = \pi_{HM}(y|x)$, as follows. If $\pi(x,y) = \pi(y)$ then $P_{x,y}(x|y) = 1 = \pi_{HM}(x|y)$. Otherwise, (x,y) is followed in the first sequence by (w,y) where $\pi(w,y) = \pi(y) > \pi(x,y)$. Provided that (w,y) does not occur also in the second sequence, $P_{x,y}(w,y) = \frac{1}{2}[\pi(y) - \pi(x,y)]$, which gives $P_{x,y}(x|y) = P_{x,y}(x,y)/[P_{x,y}(x,y) + P_{x,y}(w,y)] = 2\pi(x,y)/[\pi(y) + \pi(x,y)] = \pi_{HM}(x|y)$. If (w,y) does occur in the second sequence then it is necessary to reduce the value of $P_{x,y}(w,y)$ by $\frac{1}{2}[\pi(y) - \pi(u,v)]$, where (u,v) is the last term in the second sequence which precedes (w,y) and has $v \neq y$, to ensure that $P_{x,y}(\mathcal{X} \times \{y\}) = \frac{1}{2}[\pi(y) + \pi(x,y)]$. (The probability that is subtracted can be added to $P_{x,y}(q,r)$.) Similarly, if $\pi(x,z) = \pi(x) > \pi(x,y)$ and (x,z) occurs in the first sequence, then reduce $P_{x,y}(x,z)$ so that $P_{x,y}(\{x\} \times \mathcal{Y}) = \frac{1}{2}[\pi(x) + \pi(x,y)]$. Property (b) still holds when $P_{x,y}$ is modified in this way, because reducing $P_{x,y}(w,y)$ or $P_{x,y}(x,z)$ cannot increase $P_{x,y}(A(u,v))$ when $\pi(u,v) < 1$, and it can be verified that property (c) is satisfied.

Case (ii). $\pi(y) = 1$ and $\pi(x,y) < \beta(y)$.

(a) In this case the definition of $P_{x,y}$ is more complicated. Note first that $\eta(x) = 1$, because $\pi(x,y) < \pi(y) = 1$. Define probabilities $P_{x,y}$ on the two sequences as before. Here the first sequence must terminate at (u_2,y) since $\pi(u_2,y) = \pi(y) = 1$. Let (s,t) be the last term in the second sequence such that $t \neq y$, or take $(s,t) = (x,y)$ if there is no such term (i.e. if $\pi(x,y) = \pi(x)$). (There may be other terms (s,y) and (u_2,y) which follow (s,t) in the second sequence and have positive probability.) Let (q,r) achieve $\beta(y)$, i.e. $\pi(q,r) = \beta(y) = \pi(r)$ and $r \neq y$. Assign additional probability $\frac{1}{2}[\beta(y) - \pi(s,t)]$ to (q,r); this is non-zero only if $(q,r) \neq (s,t)$. If $\pi(q,r) < \pi(q,y)$ and $q \neq u_2$ then assign additional probability $\frac{1}{2}[\pi(q,y) - \beta(y)]$ to (q,y). (In this case $q \neq x$.) Finally, modify $P_{x,y}(u_2,y)$ so that the total probability under $P_{x,y}$ is 1.

The total probability assigned to the second sequence up to and including (s,t) is $\frac{1}{2}[\pi(s,t) + \pi(x,y)]$, and therefore $P_{x,y}(u_2,y) = 1 - \frac{1}{2}[\pi(s,t) + \pi(x,y)] - P_{x,y}(s,y) - P_{x,y}(q,r) - P_{x,y}(q,y) = 1 - \frac{1}{2}[\pi(s,t)+\pi(x,y)] - \max\{0, \frac{1}{2}[\pi(s,y) - \pi(s,t)]\} - \frac{1}{2}[\beta(y) - \pi(s,t)] - \max\{0, \frac{1}{2}[\pi(q,y) - \beta(y)]\} = 1 - \frac{1}{2}[\pi(x,y) + \max\{0, \pi(s,y) - \pi(s,t)\} + \max\{\beta(y), \pi(q,y)\}] \geq 0$, since $\pi(x,y) + \max\{0, \pi(s,y) - \pi(s,t)\} \leq \max\{\pi(s,t), \pi(s,y)\} \leq 1$ (using $\pi(x,y) \leq \pi(s,t)$) and $\max\{\beta(y), \pi(q,y)\} \leq 1$. Here $P_{x,y}(\mathcal{X}\times\{y\}) = 1-\frac{1}{2}[\pi(s,t)-\pi(x,y)]-\frac{1}{2}[\beta(y)-\pi(s,t)] = 1-\frac{1}{2}[\beta(y)-\pi(x,y)]$ and $P_{x,y}(\{x\}\times\mathcal{Y}) = \pi(x,y) + \frac{1}{2}[\pi(x) - \pi(x,y)] = \frac{1}{2}[\pi(x) + \pi(x,y)]$.

(b) We need to show that $P_{x,y}(A(w,z)) \leq \pi(w,z)$, where $A(w,z) = \{(u,v) \colon \pi(u,v) \leq \pi(w,z)\}$. As in case (i), this holds when (w,z) belongs to the second sequence, up to and including (s,t). The only other cases that need to be verified are those with $(w,z) = (q,r)$, (q,y) or (s,y), and the verification is straightforward because the probability of each term contains a factor of $\frac{1}{2}$. For example, if $\pi(q,r) < \pi(q,y) < \pi(s,y) < 1$ then $P_{x,y}(A(q,r)) = \frac{1}{2}[\beta(y) + \pi(x,y)] \leq \beta(y) = \pi(q,r)$, $P_{x,y}(A(q,y)) = \frac{1}{2}[\pi(q,y) + \pi(x,y)] \leq \pi(q,y)$, and $P_{x,y}(A(s,y)) = \frac{1}{2}[\pi(q,y) + \pi(x,y) + \pi(s,y) - \pi(s,t)] \leq \frac{1}{2}[\pi(q,y) + \pi(s,y)] \leq \pi(s,y)$.

(c) First consider conditioning on $v \in \mathcal{Y}$. When $v \neq y$, the required inequality holds as in case (i). Suppose then that $v = y$. There can be up to four values of u for which $P_{x,y}(u,y) > 0$: $u = x$, q, s or u_2. Assume that all four values are distinct. (The following argument can be modified to cover the other possible cases.) Then their probabilities are determined by $P_{x,y}(x,y) = \pi(x,y), P_{x,y}(q,y) = \max\{0, \frac{1}{2}[\pi(q,y) - \beta(y)]\}, P_{x,y}(s,y) = \max\{0, \frac{1}{2}[\pi(s,y) - \pi(s,t)]\}$, and $P_{x,y}(\mathcal{X} \times \{y\}) = 1 - \frac{1}{2}[\beta(y) - \pi(x,y)]$. Let $D = \pi(x,y) + 2 - \beta(y)$. Using Bayes' rule, $P_{x,y}(x|y) = 2\pi(x,y)/D$, $P_{x,y}(q|y) = \max\{0, [\pi(q,y) - \beta(y)]/D\}$ and $P_{x,y}(s|y) = \max\{0, [\pi(s,y) - \pi(s,t)]/D\}$. Also $\pi_{HM}(q|y) = \pi(q,y)$ when $\pi(q,y) \geq \beta(y)$, $\pi_{HM}(s|y) = 2\pi(s,y)/[\pi(s,y) + 2 - \max\{\pi(s,y), \beta(y)\}]$, $\pi_{HM}(q|y) \geq \pi_{HM}(x|y)$ and $\pi_{HM}(s|y) \geq \pi_{HM}(x|y)$, since $\pi(q,y) \geq \pi(x,y)$ and $\pi(s,y) \geq \pi(x,y)$. From these expressions it is straightforward to verify that $P_{x,y}(B|y) \leq \max\{\pi_{HM}(u|y)\colon u \in B\} = \overline{P}(B|y)$ when B is any subset of $\{x,q,s\}$. This inequality holds trivially when $u_2 \in B$ since then $\overline{P}(B|y) = \pi_{HM}(u_2|y) = 1$.

Next consider conditioning on $u \in \mathcal{X}$. By essentially the same argument as in case (i)(c), the required inequality holds for $u = q$ and for all u which appear in the second sequence, except possibly for $u = x$ and $u = u_2$. Suppose that $u = x$. Then $P_{x,y}(x,v) > 0$ only when $v = y$ or $v = v_2$, where $\pi(x,v_2) = \pi(x)$. Also $P_{x,y}(x,y) = \pi(x,y)$ and $P_{x,y}(\{x\} \times \mathcal{Y}) = \frac{1}{2}[\pi(x) + \pi(x,y)]$. By Bayes' rule, $P_{x,y}(y|x) = P_{x,y}(x,y)/P_{x,y}(\{x\} \times \mathcal{Y}) = 2\pi(x,y)/[\pi(x) + \pi(x,y)] = \pi_{HM}(y|x)$, using the fact that $\eta(x) = 1$ in case (ii). Also $P_{x,y}(v_2|x) \leq 1 = \pi_{HM}(v_2|x)$ since $\pi(x,v_2) = \pi(x)$. Thus (c) is verified for $u = x$.

Finally, suppose that $u = u_2$. Now $P_{x,y}(u_2,v) > 0$ only when $v = y$ or $v = r$ (if $q = u_2$) or $v = t$ (if $s = u_2$). It is straightforward to verify that $P_{x,y}(r|u_2) \leq \beta(y) = \pi(u_2,r) \leq \pi_{HM}(r|u_2)$ when $q = u_2$; $P_{x,y}(t|u_2) \leq \pi(u_2,t) \leq \pi_{HM}(t|u_2)$ when $s = u_2$; and $P_{x,y}(\{r,t\}|u_2) \leq \beta(y) \leq \pi_{HM}(r|u_2) \leq \max\{\pi_{HM}(r|u_2), \pi_{HM}(t|u_2)\} = \overline{P}(\{r,t\}|u_2)$ when $q = s = u_2$. This establishes (c).

(d) As shown in (c), using $\pi(y) = \eta(x) = 1$, $P_{x,y}(x|y) = 2\pi(x,y)/[\pi(x,y) + 2 - \beta(y)] = \pi_{HM}(x|y)$ and $P_{x,y}(y|x) = 2\pi(x,y)/[\pi(x) + \pi(x,y)] = \pi_{HM}(y|x)$, as required.

Case (iii). $\pi(x) = 1$ and $\pi(x,y) < \eta(x)$.
This is analogous to case (ii), with x and y interchanged.

Case (iv). $\pi(x,y) \geq \beta(y)$.

(a) Note first that $\pi(y) = 1$, because $\pi(y) \geq \pi(x,y) \geq \beta(y)$ and $\max\{\pi(y), \beta)y)\} = 1$, and $\pi(x,y) = \pi(x)$, because $\pi(x,y) \geq \beta(y) \geq \pi(x,v)$ when $v \neq y$. It follows that the two sequences defined earlier are very simple: the first sequence terminates at (u_2,y), where $\pi(u_2,y) = \pi(y) = 1$, and the second sequence terminates at (x,y). The probability measure $P_{x,y}$ is defined by $P_{x,y}(x,y) = \pi(x,y)$, $P_{x,y}(u_2,y) = 1 - \pi(x,y)$ if $u_2 \neq x$, and $P_{x,y}(u,v) = 0$ for all other $(u,v) \in \mathcal{X} \times \mathcal{Y}$.

(b) This holds because $P_{x,y}(A(w,z)) = 0$ if $\pi(w,z) < \pi(x,y)$, $P_{x,y}(A(w,z)) = \pi(x,y)$ if $\pi(x,y) \leq \pi(w,z) < 1$, and $\pi(w,z) = 1$ otherwise, and in each case $P_{x,y}(A(w,z)) \leq \pi(w,z)$.

(c) This needs to be verified only for $v = y$, $u = x$ and $u = u_2$, as only these values have positive probability under $P_{x,y}$. Here $\pi_{HM}(y|x) = \pi_{HM}(y|u_2) = \pi_{HM}(u_2|y) = 1$, since $\pi(x,y) = \pi(x)$ and $\pi(u_2,y) = \pi(u_2) = \pi(y) = 1$. Also $P_{x,y}(x|y) = \pi(x,y) = \pi_{HM}(x|y)$. Hence the required inequality holds in each case.

(d) $P_{x,y}(y|x) = 1 = \pi_{HM}(y|x)$, and $P_{x,y}(x|y) = \pi(x,y) = \pi_{HM}(x|y)$, as required.

Case (v). $\pi(x,y) \geq \eta(x)$.
This case is analogous to case (iv).

Note that cases (i)–(v) do cover all possibilities. If neither (ii) nor (iv) holds then $\pi(y) < 1$, and if neither (iii) nor (v) holds then $\pi(x) < 1$. Hence, if none of (ii)–(v) holds then $\pi(x) < 1$ and $\pi(y) < 1$, so that (i) holds. □

References

[1] B. De Baets, E. Tsiporkova, and R. Mesiar. The surprising possibilistic nature of the algebraic product. In H.-J. Zimmermann, editor, *Proceedings of EUFIT '96*, volume 1, pages 549–553, Aachen, 1996. ELITE.

[2] B. De Baets, E. Tsiporkova, and R. Mesiar. Conditioning in possibility theory with strict order norms. *Fuzzy Sets and Systems*, 1999. to appear.

[3] G. de Cooman. Describing linguistic information in a behavioural framework: possible or not? In J. Albus, A. Meystel, and R. Quintero, editors, *Intelligent Systems: A Semiotic Perspective (Proceedings of the 1996 International Multidisciplinary Conference, Gaithersburg, MD, 20–23 October 1996)*, pages 141–150. 1996.

[4] G. de Cooman. Possibility theory I: the measure- and integral-theoretic groundwork. *International Journal of General Systems*, 25:291–323, 1997.
[5] G. de Cooman. Possibility theory II: conditional possibility. *International Journal of General Systems*, 25:325–351, 1997.
[6] G. de Cooman. Possibility theory III: possibilistic independence. *International Journal of General Systems*, 25:353–371, 1997.
[7] G. de Cooman and D. Aeyels. A random set description of a possibility measure and its natural extension. 1997. submitted for publication.
[8] G. de Cooman and D. Aeyels. Supremum preserving upper probabilities. *Information Sciences*, 1998. accepted for publication.
[9] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
[10] D. Dubois, L. Fariñas Del Cerro, A. Herzig, and H. Prade. An ordinal view of independence with application to plausible reasoning. In *Uncertainty in Artificial Intelligence (Proceedings of the Tenth Conference, University of Washington, Seattle, Washington, July 29–31, 1994)*, pages 195–203.
[11] D. Dubois and H. Prade. Fuzzy logics and the generalized modus ponens revisited. *Cybernetics and Systems*, 15:293–331, 1984.
[12] D. Dubois and H. Prade. *Théorie des possibilités*. Masson, Paris, 1985.
[13] D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, New York, 1988.
[14] D. Dubois and H. Prade. The logical view of conditioning and its application to possibility and evidence theories. *International Journal of Approximate Reasoning*, 4:23–46, 1990.
[15] D. Dubois and H. Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49:65–74, 1992.
[16] E. Hisdal. Conditional possibilities independence and noninteraction. *Fuzzy Sets and Systems*, 1:283–297, 1987.
[17] H. T. Nguyen. On conditional possibility distributions. *Fuzzy Sets and Systems*, 1:299–309, 1978.
[18] A. Ramer. Conditional possibility measures. *Cybernetics and Systems*, 20:233–247, 1989.
[19] B. Schweizer and A. Sklar. Associative functions and abstract semi-groups. *Publ. Math. Debrecen*, 10:69–84, 1963.
[20] B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. Elsevier Science Publishing Company, New York, 1983.
[21] G. L. S. Shackle. *Decision, Order and Time in Human Affairs*. Cambridge University Press, Cambridge, 1961.
[22] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.
[23] N. Shilkret. Maxitive measures and integration. *Indagationes Mathematicae*, 33:109–116, 1971.
[24] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
[25] P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996.
[26] P. Walley. Statistical inferences based on a second-order possibility distribution. *International Journal of General Systems*, 26:337–383, 1997.
[27] P. Walley and G. de Cooman. A behavioural model for linguistic uncertainty. In P. P. Wang, editor, *Computing with Words*. John Wiley & Sons, New York, 1998. accepted for publication.
[28] P. M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, UK, 1975.
[29] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

Universiteit Gent, Onderzoeksgroep SYSTeMS, Technologiepark – Zwijnaarde 9, 9052 Zwijnaarde, Belgium.
E-mail address: gert.decooman@rug.ac.be