Available online at www.sciencedirect.com

SCIENCE DIRECT

Artificial Intelligence 155 (2004) 1–39

Artificial Intelligence

www.elsevier.com/locate/artint

A framework for linguistic modelling

Jonathan Lawry

Department of Engineering Mathematics, University of Bristol, Bristol BS8 1TR, UK

Received 13 November 2001; received in revised form 7 November 2003

Abstract

A new framework for linguistic reasoning is proposed based on a random set model of the degree of appropriateness of a label. Labels are assumed to be chosen from a finite predefined set of labels and the set of appropriate labels for a value is defined as a random set-valued function from a population of individuals into the set of subsets of labels. Appropriateness degrees are then evaluated relative to the distribution on this random set where the appropriateness degree of a label corresponds to the probability that it is contained in the set of appropriate labels. This interpretation is referred to as label semantics. A natural calculus for appropriateness degrees is described which is weakly functional while taking into account the logical structure of expressions. Given this framework it is shown that a bayesian approach can be adopted in order to infer probability distributions on the underlying variable given constraints both in the form of linguistic expressions and mass assignments. In addition, two conditional measures are introduced for evaluating the appropriateness of a linguistic expression given other linguistic information.

Keywords: Random sets; Linguistic constraints; Fuzzy labels; Label semantics; Bayesian inference

1. Introduction

The limitations of classical modelling techniques to effectively capture the behaviour of complex systems has become increasingly clear over recent years. This has motivated research into new, alternative modelling paradigms by the artificial intelligence community (e.g., fuzzy reasoning, possibility theory, Bayesian modelling, default reasoning: see [4,8, 11,28,30]). All of these approaches share an emphasis on high level qualitative descriptions as opposed to a more traditional low level framework. The advantage of such higher-level knowledge representation is that it allows for the fusion of expert or background knowl-

E-mail address: j.lawry@bris.ac.uk (J. Lawry).

0004-3702/$ – see front matter
doi:10.1016/j.artint.2003.10.001

edge and knowledge derived from data. Furthermore, it tends to provide a clearer insight into the underlying nature of the system than can be obtained from less transparent lower-level models. Another feature shared by many of the new approaches is that they provide a methodology for reasoning in the presence of uncertainty. This is no coincidence, but rather is due to the fact that uncertainty and imprecision are often inherent in complex modelling problems. This uncertainty is not only due to lack of precision or errors in measured features but is often present in the model itself since the available features may not be sufficient to provide a complete model of the system. To illustrate this point, consider the important area of river basin modelling for flood forecasting. For this problem it is often necessary to model river levels at a particular time point, purely in terms of rainfall and river levels at earlier times. However, in reality so many complex features influence runoff that it is both difficult to identify the most important and practically impossible to measure any but a few of them. For instance, the likelihood that a given rainfall event will produce a flood is dramatically affected by such factors as the size of the drainage basin, the topography of the basin, the amount of urban use within the basin and so on.

While the development of analytical models may be impractical for many complex systems, there is often data available implicitly describing the behaviour of the system. For example, large companies such as supermarkets, high street stores and banks collect a stream of data relating to the behaviour of their customers. Such data must be analysed to provide flexible models of customer behaviour that can be used to aid a wide variety of decision-making processes. Hence, if a higher level modelling approach is to be truly effective it must provide a natural knowledge representation framework for inductive learning. As such it is important that it allows for the modelling of uncertainty, imprecision and vagueness in a semantically clear manner. Indeed, we should emphasise the necessity of a clear underlying semantics for any higher-level modelling paradigm since one of the fundamental reasons for a high level approach is to provide transparent models that can be understood and used by practitioners in the relevant fields. This cannot be achieved if the validity of the underlying concepts and inference processes are either obscured or in doubt. In the sequel we will outline a new methodology for linguistic modelling and show how it can be applied in an inductive learning context. The approach will centre on the modelling of linguistic constraints on variables as proposed by Zadeh [37] although the underlying semantics will be quite different.

The phrase computing with words was introduced by Zadeh [42] to capture the idea of computation based not on numerical values, but on natural language terms and expressions. As a general idea this is clearly of relevance to the type of modelling described above, however, we shall propose a quite different interpretation to that given in [38–40]. The general methodology for computing with words proposed by Zadeh is that of fuzzy set theory or fuzzy logic and in particular is based on the idea of linguistic variables (see [38–40]). A linguistic variable is defined as a variable that takes natural language terms such as *large, small, tall, medium* etc. as values and where the meaning of these words is given by fuzzy sets on some underlying domain of discourse. Hence, a particular expression of the form *Bill is tall* can be taken as expressing the fact that the linguistic variable describing Bill's height has the value tall, and such a statement has a partial truth-value corresponding to the membership degree of Bill's actual height in the fuzzy set representing the meaning of tall. The truth-value of compound expressions such as *Bill is tall or medium* is then evaluated according to a fuzzy set calculus based on some choice of t-norm or t-conorm (see [18] for an exposition).

In our view the principal problem with the above approach is that the semantics underlying standard fuzzy logic or indeed the notion of membership function itself is rather obscure. The difficulty is revealed by consideration of a fundamental question that should be asked of all models of linguistic constraints. What information is conveyed regarding the underlying variable? For instance, if someone asserts that *Bill is tall* exactly what information about Bill's height is conveyed by that assertion? In the case of fuzzy set theory, according to Zadeh [41], the latter provides a flexible constraint on the variable representing Bill's height. More specifically, it tells us that the possibility distribution on Bill's height corresponds to the membership function of the fuzzy set *tall*. However, this association with possibility distributions does not, in itself, support the assumption of a fully truth-functional calculus for membership degrees, as in fuzzy set theory (see [26]). Indeed, it does not really provide any insight into the behaviour of compound fuzzy sets. One possible solution to this difficulty is to accept that neither possibility distributions or fuzzy memberships are sufficiently intuitive to be treated as primitive notions and attempt to provide a lower-level model. If we are going to adopt the fuzzy logic methodology then any such semantics should not only be intuitive but should also be consistent with a fully truth-functional calculus based on a particular choice of t-norm and t-conorm. A number of different models have been investigated and these are reviewed in [6,27].

One of the most promising ideas is to view fuzzy memberships as being fixed point coverage functions of random sets, themselves representing uncertainty or variation in the underlying crisp definition of a concept [12]. For instance, we might have a population of different individuals each proposing their own set of heights that would qualify for the description *tall*. The associated random set would be a function from the set of individuals into the set of subsets of heights and the membership of a particular height, h, in the fuzzy set *tall* would correspond to the probability of encountering an individual who included h in their crisp set definition. This is the essence of the voting model for fuzzy sets proposed originally by Black [3] and later by Gaines [10]. Clearly this interpretation is implicitly probabilistic in its nature and hence, it is not perhaps surprising that it does not fit well within the inference framework of fuzzy logic. One problem is that there is not a one-to-one correspondence between fuzzy sets and random sets. The same fuzzy set could be generated by a potentially infinite family of random sets (see Goodman [12]). In possibility theory this problem is overcome by making the assumption that the random set is consonant (i.e., the set of sets with non-zero mass constitutes a nested hierarchy). Lawry [20] justifies this by introducing the idea of an optimism parameter according to which the more optimistic a voter the more likely they are to include h in the extension of the concept *tall*. It is difficult to consolidate such an assumption with a fully truth-functional calculus since, in that case, a voter with a high optimism parameter would be required to be optimistic regarding both the concept and its negation. Lawry [20] suggests a weaker notion of negation to overcome this difficulty but nonetheless the treatment of negation remains problematic. In the sequel we outline a new random set based approach but where the random sets relate to sets of appropriate labels for a value. We describe a formal framework for such an approach and show how it overcomes some of the problems highlighted above. This new calculus, however, will not be fully truth-functional but rather functional in a weaker, although sufficient, sense. The work described in Sections 2 and 3 is an extension of that presented in an earlier paper [22]. This work is clearly related to random set semantics for fuzzy sets as proposed by Goodman [12] and Nguyen [25]. However, the latter defines random sets on the underlying attribute universe whereas our proposed framework will define random sets over labels. In our view the focus on the labels themselves provides an interesting new perspective.

2. Label semantics

The fundamental notion underlying label semantics is that when individuals make assertions of the kind described above they are essentially providing information about what labels are appropriate for the value of some underlying variable. For simplicity, we assume that for a given context only a finite set of words is available. This is a somewhat controversial assumption since it might be claimed that by recursively applying hedges we can easily generate an infinite set of labels from an initially finite set of words. In other words, if *tall* is a possible label for Bill's height then so is *very tall*, *quite tall*, *very very tall* and so on. This claim is problematic, however, for a number of reasons. For instance, it would appear that the use of hedges in natural language is somewhat restricted. One might use the expressions *very tall* and *quite tall* but *very quite tall* or even *quite very tall* are never used. Also, there seems in practice to be a limit on the number of times hedges can be applied to a label before it becomes nonsensical. This latter point seems to suggest that in practice only a finite number of labels may be available even in natural language. Another related difficulty with the use of hedges is determining the relationship between the meaning of a word and the meaning of any new word generated from it by application of some hedge. In Zadeh [38–40] it is suggested that such relationships have a simple functional form. For example, if the meaning of *tall* is defined by a fuzzy set with membership function μ_{tall} then Zadeh proposes that the meaning of *very tall* is the fuzzy set with membership μ_{tall}^2. The choice of this particular function seems relatively arbitrary and indeed, perhaps more fundamentally, it is far from apparent that there should be any such simple functional relationship between the meaning of a word and that of a new word generated by application of a hedge. In other words, we would claim that while hedges are a simple syntactic device for generating new labels there is no equally simple semantic device for generating the associated new meanings.

Now let us return to the problem of interpreting natural language statements regarding, say, Bill's height as represented by variable H. Let us suppose then that there is a fixed finite set of possible labels for H, denoted *LA*, and that these labels are both known and completely identical for any individual who will make or interpret a statement regarding Bill's height. Given these assumptions how can we now interpret a statement such as *Bill is tall* as asserted by a particular individual I? We claim that one natural interpretation is that it merely conveys the information that, according to I, *tall* is an appropriate label for the value of H. In order to clarify this idea suppose I knows that $H = h$ and that given this information he/she is able to identify a subset of *LA* consisting of those words appropriate as labels for the value h. This set is denoted $\mathcal{D}_h^I$ which stands for the description of h given by I. If we allow I to vary across some population of individuals V then we naturally obtain a random set $\mathcal{D}_h$ from V into the power set of LA such that $\mathcal{D}_h(I) = \mathcal{D}_h^I$. Given this we can obtain higher level information about the degree of applicability of a label to a value by defining, in this case, $\mu_{tall}(h) = Pr(\{I \in V \mid tall \in \mathcal{D}_h^I\})$ where the latter probability is calculated on the basis of some underlying prior distribution on V. Now clearly this is a function from Ω into $[0, 1]$ and therefore can technically be viewed as a fuzzy set. However, we shall use the term 'appropriateness degree' partly because it more accurately reflects the underlying semantics and partly to highlight the quite distinct calculus for these functions that will be introduced in the sequel. Similarly we can determine a probability distribution (or mass assignment) for the random set $\mathcal{D}_h$ by defining $\forall S \subseteq LA$ $m_h(S) = Pr(\{I \in V \mid \mathcal{D}_h^I = S\})$. Now suppose that I does not know the value of H (or alternatively we do not know the value assigned to H by I) then they (we) would naturally define a random set $\mathcal{D}_H^I$ from the universe of H into the power set of LA such that $\mathcal{D}_H^I(h) = \mathcal{D}_h^I$. The distribution of this random set will clearly depend on the prior information available regarding the distribution of H. Hence, the assertion by I that *Bill is tall* would in this context be interpreted as $tall \in \mathcal{D}_H^I$. Finally in the case when we have no information regarding I then we can define a random set $\mathcal{D}_H$ from the cross product of V and the universe of H into the power set of LA such that $\mathcal{D}_H(I, h) = \mathcal{D}_h^I$ and interpret the above statement as $tall \in \mathcal{D}_H$. In order to clarify some of these ideas consider the following example where the objective is to provide linguistic labels for the outcome of a single throw of a dice.

Example 1. Suppose the variable *SCORE* with universe $\{1, 2, 3, 4, 5, 6\}$ gives the outcome of a single throw of a particular dice. Let $LA = \{low, medium, high\}$ and $V = \{I_1, I_2, I_3\}$ then a possible definition of $\mathcal{D}_{SCORE}$ is as follows:

$$\begin{aligned}
&\mathcal{D}_1^{I_1} = \mathcal{D}_1^{I_2} = \mathcal{D}_1^{I_3} = \{low\}, \quad \mathcal{D}_2^{I_1} = \{low, medium\}, \quad \mathcal{D}_2^{I_2} = \{low\}, \quad \mathcal{D}_2^{I_3} = \{low\},\\
&\mathcal{D}_3^{I_1} = \{medium\}, \quad \mathcal{D}_3^{I_2} = \{medium\}, \quad \mathcal{D}_3^{I_3} = \{medium, low\},\\
&\mathcal{D}_4^{I_1} = \{medium, high\}, \quad \mathcal{D}_4^{I_2} = \{medium\}, \quad \mathcal{D}_4^{I_2} = \{medium\},\\
&\mathcal{D}_5^{I_1} = \{high\}, \quad \mathcal{D}_5^{I_2} = \{medium, high\}, \quad \mathcal{D}_5^{I_3} = \{high\},\\
&\mathcal{D}_6^{I_1} = \mathcal{D}_6^{I_2} = \mathcal{D}_6^{I_3} = \{high\}.
\end{aligned}$$

The value of the appropriateness measure will depend on the underlying distribution on $V = \{I_1, I_2, I_3\}$, perhaps representing the weight of importance associated with the views of each individual. For instance, if we assume a uniform distribution on V then the degree of appropriateness of *low* as a label for 3 is given by

$$\frac{|\{I \in V \mid low \in \mathcal{D}_3^I\}|}{|V|} = \frac{|\{I_3\}|}{|V|} = \frac{1}{3}.$$

Overall the appropriateness degrees for each word are given by

$$\begin{aligned}
&\mu_{low}(1) = \mu_{low}(2) = 1, \quad \mu_{low}(3) = \frac{1}{3},\\
&\mu_{medium}(2) = \frac{1}{3}, \quad \mu_{medium}(3) = 1, \quad \mu_{medium}(4) = 1, \quad \mu_{medium}(5) = \frac{1}{3},\\
&\mu_{high}(4) = \frac{1}{3}, \quad \mu_{high}(5) = 1, \quad \mu_{high}(6) = 1.
\end{aligned}$$

Similarly, assuming a uniform prior on V we can determine mass assignments on $\mathcal{D}_{SCORE}$ for $SCORE = 1, \ldots, 6$. For example, if $SCORE = 2$ we have

$$m_2\big(\{low, medium\}\big) = \frac{|\{I \in V \mid \mathcal{D}_2^I = \{low, medium\}\}|}{|V|} = \frac{|\{I_1\}|}{|V|} = \frac{1}{3}.$$

The mass assignments for each value of x are given by

$$\begin{aligned}
&m_1 = \{low\} : 1, \quad m_2 = \{low, medium\} : \frac{1}{3},\ \{low\} : \frac{2}{3},\\
&m_3 = \{medium\} : \frac{2}{3},\ \{low, medium\} : \frac{1}{3},\\
&m_4 = \{medium, high\} : \frac{1}{3},\ \{medium\} : \frac{2}{3},\\
&m_5 = \{high\} : \frac{2}{3},\ \{medium, high\} : \frac{1}{3}, \quad m_6 = \{high\} : 1.
\end{aligned}$$

In order, to determine an overall mass assignment m as $SCORE$ varies, we need to know the distribution on the universe $\{1, \ldots, 6\}$. Assuming a uniform distribution gives, for example,

$$m\big(\{low, medium\}\big) = \sum_{x=1}^{6} m_x\big(\{low, medium\}\big) Pr(x) = \frac{\frac{1}{3} + \frac{1}{3}}{6} = \frac{1}{9}.$$

Overall we have:

$$\begin{aligned}
m = {} & \{low\} : \frac{5}{18},\ \{low, medium\} : \frac{1}{9},\\
& \{medium\} : \frac{2}{9},\ \{medium, high\} : \frac{1}{9},\ \{high\} : \frac{5}{18}.
\end{aligned}$$

We now consider the problem of how to interpret expressions involving compound labels built up using some set of logical connectives. For the scope of this paper we will consider the four main connectives $\wedge$, $\vee$, $\rightarrow$ and $\neg$. Firstly, let us consider the case of negation. How do we interpret expressions of the form *Bill is not tall*? We take the view here that negation is used to express the non-suitability of a label. In other words the above statement means that *tall* is not an appropriate label for H, or $tall \notin \mathcal{D}_H$. Conjunction and disjunction are then taken as having the obvious meanings so that *Bill is tall and medium* is interpreted as saying that both *tall* and *medium* are appropriate as labels for H (i.e., $\{tall, medium\} \subseteq \mathcal{D}_H$), and *Bill is tall or medium* is interpreted as saying that either *tall* is an appropriate label for H or *medium* is an appropriate label for H (i.e., $tall \in \mathcal{D}_H$ or $medium \in \mathcal{D}_H$). In the case of implication we take *very tall implies tall* to mean that whenever *very tall* is an appropriate label for H so is *tall* (i.e., $very\ tall \in \mathcal{D}_H$ implies that $tall \in \mathcal{D}_H$).

It will clearly be desirable to measure the appropriateness degree of such compound expressions for particular values of the underlying variable. For instance, given the scenario outlined in Example 1 we might want to know what is the appropriateness degree of the expression *medium or low* to the value 3. Now this expression identifies the set of subsets of LA which either contains *low* or *medium* (i.e., $\{\{low\}, \{medium\},$

$\{low, medium\}$, $\{low, high\}$, $\{medium, high\}$, $\{low, medium, high\}\})$. Hence, it is natural to define the appropriateness degree of *medium or low* to 3 as the sum of the values of m_3 across this set (i.e., $m_3(\{low\}) + m_3(\{medium\}) + m_3(\{low, medium\}) + m_3(\{low, high\}) + m_3(\{medium, high\}) + m_3(\{low, medium, high\}) = \frac{2}{3} + \frac{1}{3} = 1$). In the following section we formalise the above ideas within a logical framework.

3. A formal framework for label semantics

In this paper we adopt a logical formalisation for label semantics where label expressions are represented by propositional logic sentences. Consider a formal language consisting of the set of the labels $LA = \{L_1, \ldots, L_n\}$ and the connectives $\wedge$, $\vee$, $\rightarrow$ and $\neg$. Within this language we can represent compound linguistic descriptions generated recursively by application of the connectives:

Definition 2 (*Label expressions*). The set of label expressions, *LE*, is defined recursively as follows:

(i) $L_i \in LE$ for $i = 1, \ldots, n$.
(ii) If $\theta, \varphi \in LE$ then $\neg\theta, \theta \wedge \varphi, \theta \vee \varphi, \theta \rightarrow \varphi \in LE$.

Recall from the discussion in the previous section that a label expression identifies a set of subsets of *LA* which capture its meaning. We now give a formal definition of this subset for any general label expression:

Definition 3 (*Appropriate label sets*). Every $\theta \in LE$ is associated with a set of subsets of *LA* (i.e., an element of $2^{2^{LA}}$), denoted $\lambda(\theta)$ and defined recursively as follows:

(i) $\lambda(L_i) = \{S \subseteq LA \mid L_i \in S\}$.
(ii) $\lambda(\theta \wedge \varphi) = \lambda(\theta) \cap \lambda(\varphi)$.
(iii) $\lambda(\theta \vee \varphi) = \lambda(\theta) \cup \lambda(\varphi)$.
(iv) $\lambda(\theta \rightarrow \varphi) = \overline{\lambda(\theta)} \cup \lambda(\varphi)$.
(v) $\lambda(\neg\theta) = \overline{\lambda(\theta)}$.

Intuitively $\lambda(\theta)$ corresponds to those subsets of *LA* identified as being candidates for the set of appropriate labels for x (i.e., possible values for $\mathcal{D}_x$) by expression θ. In this sense the imprecise linguistic restriction 'x is θ' on x corresponds to the strict constraint $\mathcal{D}_x \in \lambda(\theta)$ on $\mathcal{D}_x$. Hence, the linguistic description $\mathcal{D}_x$ can provide an alternative to linguistic variables (Zadeh [38–40]) as a means of formally representing linguistic constraints.

Example 4. Let $LA = \{small, medium, large\}$ then

$$\lambda(small \wedge medium) = \big\{\{small, medium\},\ \{small, medium, large\}\big\},$$
$$\lambda(small \vee medium) = \big\{\{small\},\ \{medium\},\ \{small, medium\},\ \{small, large\},$$
$$\{medium, large\},\ \{small, medium, large\}\big\},$$

$$\lambda(small \to medium) = \big\{\{small, medium\},\ \{small, medium, large\},\ \{medium, large\},\ \{medium\},\ \{large\},\ \emptyset\big\},$$
$$\lambda(\neg small) = \big\{\{medium\},\ \{large\},\ \{medium, large\},\ \emptyset\big\}.$$

The following results illustrate the clear relationship between appropriate label sets and the logical structure of the expressions that identify them. Initially, however, we introduce some basic notation. Let *Val* denote the set of valuations (i.e., allocations of truth values) on $\{L_1, \ldots, L_n\}$. For $v \in Val$ $v(L_i) = true$ can be taken as meaning that L_i is an appropriate label in the current context. Let $LE^0 = \{L_1, \ldots, L_n\}$ and $LE^{n+1} = LE^n \cup \{\neg\theta, \theta \wedge \varphi, \theta \vee \varphi, \theta \to \varphi \mid \theta, \varphi \in LE^n\}$. Clearly we have that $LE = \bigcup_n LE^n$ and also, from a valuation v on LE^0 the truth-value, $v(\theta)$, for $\theta \in LE$ can be determined recursively in the usual way by application of the truth tables for the connectives.

Definition 5. Let $\tau : Val \to 2^{LA}$ such that $\forall v \in Val$ $\tau(v) = \{L_i \mid v(L_i) = true\}$.

Notice that τ is clearly a bijection. Also note that for $v \in Val$ $\tau(v)$ can be associated with a Herbrand interpretation of the language *LE* (see [24]).

Lemma 6. $\forall \theta \in LE$ $\{\tau(v) \mid v \in Val, v(\theta) = true\} = \lambda(\theta)$.

Proof. We prove this by induction on the complexity of θ.

Suppose $\theta \in LE^0$, so that $\theta = L_i$ for some $i \in \{1, \ldots, n\}$. Now as v ranges across all valuations for which L_i is true, then $\tau(v)$ ranges across all subsets of *LA* that contain L_i. Hence, $\{\tau(v) \mid v \in Val, v(L_i) = true\} = \{S \subseteq LA \mid \{L_i\} \subseteq S\} = \lambda(L_i)$ as required.

Now suppose we have $\forall \theta \in LE^n$, $\{\tau(v) \mid v \in Val, v(\theta) = true\} = \lambda(\theta)$ and consider an expression $\theta \in LE^{n+1}$ then either $\theta \in LE^n$ in which case the result follows trivially or one of the following hold:

(i) $\theta = \phi \wedge \varphi$ where $\phi, \varphi \in LE^n$. In this case

$$\begin{aligned}&\big\{v \in Val \mid v(\phi \wedge \varphi) = true\big\}\\ &\quad = \big\{v \in Val \mid v(\phi) = true\big\} \cap \big\{v \in Val \mid v(\varphi) = true\big\}.\end{aligned}$$

Therefore,

$$\begin{aligned}&\big\{\tau(v) \mid v \in Val, v(\phi \wedge \varphi) = true\big\}\\ &\quad = \big\{\tau(v) \mid v \in Val, v(\phi) = true\big\} \cap \big\{\tau(v) \mid v \in Val, v(\varphi) = true\big\}\\ &\quad = \lambda(\phi) \cap \lambda(\varphi) \quad \text{(by the inductive hypothesis)}\\ &\quad = \lambda(\phi \wedge \varphi) \quad \text{by Definition 3.}\end{aligned}$$

(ii) $\theta = \phi \vee \varphi$ where $\phi, \varphi \in LE^n$. In this case

$$\big\{v \in Val \mid v(\phi \vee \varphi) = true\big\} = \big\{v \in Val \mid v(\phi) = true\big\} \cup \big\{v \in Val \mid v(\varphi) = true\big\}.$$

Therefore,

$$\begin{aligned}&\{\tau(v) \mid v \in Val, v(\phi \vee \varphi) = true\}\\&\quad = \{\tau(v) \mid v \in Val, v(\phi) = true\} \cup \{\tau(v) \mid v \in Val, v(\varphi) = true\}\\&\quad = \lambda(\phi) \cup \lambda(\varphi) \quad \text{(by the inductive hypothesis)}\\&\quad = \lambda(\phi \vee \varphi) \quad \text{by Definition 3.}\end{aligned}$$

(iii) $\theta = \phi \to \varphi$ where $\phi, \varphi \in LE^n$. In this case

$$\begin{aligned}&\{v \in Val \mid v(\phi \to \varphi) = true\}\\&\quad = \{v \in Val \mid v(\phi) = false\} \cup \{v \in Val \mid v(\varphi) = true\}\\&\quad = \overline{\{v \in Val \mid v(\phi) = true\}} \cup \{v \in Val \mid v(\varphi) = true\}.\end{aligned}$$

Therefore,

$$\begin{aligned}&\{\tau(v) \mid v \in Val, v(\phi \to \varphi) = true\}\\&\quad = \overline{\{\tau(v) \mid v \in Val, v(\phi) = true\}} \cup \{\tau(v) \mid v \in Val, v(\varphi) = true\}\\&\quad = \overline{\lambda(\phi)} \cup \lambda(\varphi) \quad \text{(by the inductive hypothesis)}\\&\quad = \lambda(\phi \to \varphi) \quad \text{by Definition 3.}\end{aligned}$$

(iv) $\theta = \neg\phi$ where $\phi \in LE^n$. In this case

$$\begin{aligned}&\{\tau(v) \mid v \in Val, v(\neg\phi) = true\}\\&\quad = \overline{\{\tau(v) \mid v \in Val, v(\phi) = true\}}\\&\quad = \overline{\lambda(\phi)} \quad \text{(by the inductive hypothesis)}\\&\quad = \lambda(\neg\phi) \quad \text{by Definition 3.} \qquad \square\end{aligned}$$

Proposition 7. *For $\theta, \varphi \in LE$ $\theta \models \varphi$ iff $\lambda(\theta) \subseteq \lambda(\varphi)$.*

Proof. ($\Rightarrow$)

$$\begin{aligned}\theta \models \varphi &\Rightarrow \{v \in Val \mid v(\theta) = true\} \subseteq \{v \in Val \mid v(\varphi) = true\}\\&\Rightarrow \{\tau(v) \mid v \in Val, v(\theta) = true\} \subseteq \{\tau(v) \mid v \in Val, v(\varphi) = true\}\\&\Rightarrow \lambda(\theta) \subseteq \lambda(\varphi) \quad \text{by Lemma 6.}\end{aligned}$$

($\Leftarrow$) Suppose $\lambda(\theta) \subseteq \lambda(\varphi)$. Then $\lambda(\theta) = \{\tau(v) \mid v \in Val, v(\theta) = true\}$ and $\lambda(\varphi) = \{\tau(v) \mid v \in Val, v(\varphi) = true\}$ by Lemma 6.

Therefore

$$\begin{aligned}&\{\tau(v) \mid v \in Val, v(\theta) = true\} \subseteq \{\tau(v) \mid v \in Val, v(\varphi) = true\}\\&\quad \Rightarrow \{v \in Val \mid v(\theta) = true\} \subseteq \{v \in Val \mid v(\varphi) = true\}\end{aligned}$$

since τ is a bijection. $\square$

A trivial corollary of this proposition is:

Corollary 8. *For $\theta, \varphi \in LE$ $\theta \equiv \varphi$ iff $\lambda(\theta) = \lambda(\varphi)$.*

Proposition 9. *If $\varphi \in LE$ is inconsistent then $\lambda(\varphi) = \emptyset$.*

Proof. If $\varphi \in LE$ is inconsistent then $\varphi \equiv \theta \wedge \neg\theta$ so that by Corollary 8. $\lambda(\varphi) = \lambda(\theta \wedge \neg\theta) = \lambda(\theta) \cap \lambda(\neg\theta) = \lambda(\theta) \cap \overline{\lambda(\theta)} = \emptyset$ by Definition 3. □

In order to introduce higher level measures of appropriateness as discussed in earlier sections we need to consider the logical structure of label expressions in conjunction with a set of individuals, a probability distribution on that set and a probability distribution on the domain Ω. To allow for this we now introduce the notions of a frame and an extended frame.

Definition 10 (*Frame and extended frame*).

(i) A frame is a tuple $\langle V, P_V \rangle$ where V is a set of individuals and P_V is a probability distribution on V.
(ii) An extended frame is a tuple $\langle V, P_V, \Omega, P_\Omega \rangle$ where $\langle V, P_V \rangle$ is a frame and P_Ω is a distribution on the underlying domain Ω.

Definition 11 (*Mass assignments and label appropriateness degrees*).

(i) Given a frame $\Gamma = \langle V, P_V \rangle$ we define the mass assignment of $\mathcal{D}_x$ by

$$\forall S \subseteq LA \quad m_x^\Gamma(S) = P_V\big(\big\{I \in V \mid \mathcal{D}_x^I = S\big\}\big).$$

In the case where V is finite and P_V is the uniform distribution this corresponds to

$$\forall S \subseteq LA \quad m_x^\Gamma(S) = \frac{|\{I \in V \mid \mathcal{D}_x^I = S\}|}{|V|}.$$

From this mass assignment we define the appropriateness measure (or degree) μ^Γ by

$$\forall \theta \in LE,\ \forall x \in \Omega \quad \mu_\theta^\Gamma(x) = \sum_{S \in \lambda(\theta)} m_x^\Gamma(S).$$

(ii) Given an extended frame $\Gamma^+ = \langle V, P_V, \Omega, P_\Omega \rangle$ then the mass assignment of $\mathcal{D}_x$ as x varies across Ω is given by

$$\forall S \subseteq LA \quad m^{\Gamma^+}(S) = P_V \times P_\Omega\big(\big\{\langle I, x \rangle \mid \mathcal{D}_x^I = S\big\}\big).$$

From this mass assignment we define the general appropriateness degree by

$$\forall \theta \in LE \quad \mu_\theta^{\Gamma^+} = \sum_{S \in \lambda(\theta)} m^{\Gamma^+}(S).$$

In cases where the frame or extended frame is fixed we shall drop the superscripts Γ and Γ^+ in the above definitions.

Notice that it is not a requirement of Definition 11 that zero mass be allocated to the empty set. In label semantics $m_x^\Gamma(\emptyset)$ corresponds to the probability in frame

Γ that no labels are appropriate to describe x (i.e., that $\mathcal{D}_x = \emptyset$). In terms of appropriateness degrees, allocating a non-zero value to $m_x^{\Gamma}(\emptyset)$ has the consequence that $\max(\mu_{L_1}^{\Gamma}(x), \ldots, \mu_{L_n}^{\Gamma}(x)) < 1$.

Trivially, by Proposition 7 we have that if $\theta \models \varphi$ then for any frame Γ $\forall x \in \Omega$ $\mu_{\theta}^{\Gamma}(x) \leqslant \mu_{\varphi}^{\Gamma}(x)$ and for any extended frame Γ^+ $\mu_{\theta}^{\Gamma^+} \leqslant \mu_{\varphi}^{\Gamma^+}$. Similarly by Corollary 8 we have that if $\theta \equiv \varphi$ then for any frame Γ, $\forall x \in \Omega$ $\mu_{\theta}^{\Gamma}(x) = \mu_{\varphi}^{\Gamma}(x)$ and for any extended frame Γ^+ $\mu_{\theta}^{\Gamma^+} = \mu_{\varphi}^{\Gamma^+}$.

Proposition 12. *For any frame Γ, $\forall \theta \in LE$, $\forall x \in \Omega$ $\mu_{\neg\theta}^{\Gamma}(x) = 1 - \mu_{\theta}^{\Gamma}(x)$.*

Proof.

$$\mu_{\neg\theta}^{\Gamma}(x) = \sum_{S \in \lambda(\neg\theta)} m_x^{\Gamma}(S) = \sum_{S \in \overline{\lambda(\theta)}} m_x^{\Gamma}(S) = 1 - \sum_{S \in \lambda(\theta)} m_x^{\Gamma}(S) = 1 - \mu_{\theta}^{\Gamma}(x). \qquad \square$$

In order to investigate the behaviour of the appropriateness measure on conjunctions, disjunctions and implications we need to introduce the notion of consonant mass assignments. More specifically, we will only consider frames Γ such that m_x^{Γ} is consonant for all $x \in \Omega$. Here, consonance has the standard random set meaning (see [13]) that $\forall S, S' \subseteq LA$ if both $m_x^{\Gamma}(S) > 0$ and $m_x^{\Gamma}(S') > 0$ then either $S \subseteq S'$ for $S' \subseteq S$.

Consonance of label sets implies that individuals in V differ regarding what labels are appropriate for a value only in terms of generality or specificity. This could be justified by the idea that all individuals share a common ordering on the appropriateness of labels for a value and that the composition of $\mathcal{D}_x^I$ is consistent with this ordering for each I. More formally, supposing for each element $x \in \Omega$ the population V shares a common total ordering $\preceq_x$ where for $L_i, L_j \in LA$, $L_i \preceq_x L_j$ means that L_j is as least as appropriate as a label for x as L_i. In this case, when deciding on a set of appropriate labels, an individual I would be expected to be consistent with $\preceq_x$ so that if $L_i \in \mathcal{D}_x^I$ then L_j will also be in $\mathcal{D}_x^I$ for all labels L_j such that $L_i \preceq_x L_j$. Clearly, in this case as we vary individuals across V then the values of $\mathcal{D}_x^I$ occurring will form a nested hierarchy. For instance, in the case of the dice problem described in Example 1 possible appropriateness orderings for values $SCORE = 1, \ldots, 6$ are as follows:

$$\begin{aligned}
&high \preceq_1 medium \preceq_1 low,\ \ high \preceq_2 medium \preceq_2 low,\\
&high \preceq_3 low \preceq_3 medium,\ \ low \preceq_4 high \preceq_4 medium,\\
&low \preceq_5 medium \preceq_5 high,\ \ low \preceq_6 medium \preceq_6 high.
\end{aligned}$$

Hence, for any individual I, if I decides that *low* is an appropriate label for 3 ($low \in \mathcal{D}_3^I$) then to be consistent with the ordering $\preceq_3$ they must also decide that *medium* is an appropriate label for 3 ($medium \in \mathcal{D}_3^I$) since *medium* is at least as appropriate as *low* as a label for 3.

Notice, that the consonance assumption for random sets on labels is in one sense weaker than the corresponding assumption for random sets on the universe Ω, since the latter requires individuals to maintain the same level of specificity across all values in Ω. To see this more clearly recall Example 1 and observe that m_x is consonant $\forall x \in \{1, \ldots, 6\}$. Now for each member $I \in V$ the extension (associated subset of Ω) of, say *medium* is given by $\{x \in \Omega \mid medium \in \mathcal{D}_x^I\}$. Hence, we obtain $\{2,3,4\}$, $\{3,4,5\}$ and $\{3,4\}$ for I_1, I_2 and I_3 respectively. Clearly, however, this does not form a nested hierarchy.

Proposition 13. *If $\forall x \in \Omega$ m_x^Γ is a consonant mass assignment then for $L_i, L_j \in LA$ we have that $\forall x \in \Omega$ $\mu_{L_i \wedge L_j}^\Gamma(x) = \min(\mu_{L_i}^\Gamma(x), \mu_{L_j}^\Gamma(x))$.*

Proof. Notice

$$\begin{aligned}\lambda(L_i \wedge L_j) &= \lambda(L_i) \cap \lambda(L_j) = \bigl\{S \subseteq LA \mid \{L_i\} \subseteq S\bigr\} \cap \bigl\{S \subseteq LA \mid \{L_j\} \subseteq S\bigr\}\\ &= \bigl\{S \subseteq LA \mid \{L_i, L_j\} \subseteq S\bigr\}.\end{aligned}$$

Hence,

$$\forall x \in \Omega \quad \mu_{L_i \wedge L_j}^\Gamma(x) = \sum_{S:\ \{L_i, L_j\} \subseteq S} m_x^\Gamma(S).$$

For any x, since m_x^Γ is a consonant mass assignment then it must have the form $m_x^\Gamma = M_0 : m_0, \ldots, M_k : m_k$ where $M_t \subset M_{t+1}$ for $t = 0, \ldots, k-1$. Now suppose w.l.o.g. that $\mu_{L_i}^\Gamma(x) \leqslant \mu_{L_j}^\Gamma(x)$ then $\{L_i\} \subseteq M_t$ iff $\{L_i, L_j\} \subseteq M_t$ for $t = 0, \ldots, k$. Therefore

$$\mu_{L_i \wedge L_j}^\Gamma(x) = \sum_{S:\ \{L_i\} \subseteq S} m_x^\Gamma(S) = \mu_{L_i}^\Gamma(x) = \min\bigl(\mu_{L_i}^\Gamma(x), \mu_{L_j}^\Gamma(x)\bigr). \qquad \Box$$

Proposition 14. *If for all $x \in \Omega$ m_x^Γ is a consonant mass assignment then for $L_i, L_j \in LA$ we have that $\forall x \in \Omega$ $\mu_{L_i \vee L_j}^\Gamma(x) = \max(\mu_{L_i}^\Gamma(x), \mu_{L_j}^\Gamma(x))$.*

In order to compare and contrast label semantics with the many-valued logic approach to fuzzy reasoning we first give a formal definition of what is meant for a calculus to be strongly functional.

Definition 15. Let $w : LE \times \Omega \to [0,1]$ then w is said to be strongly functional iff there exist functions $f_\neg : [0,1] \to [0,1]$, $f_\wedge : [0,1]^2 \to [0,1]$, $f_\vee : [0,1]^2 \to [0,1]$ and $f_\to : [0,1]^2 \to [0,1]$ such that $\forall \theta, \varphi \in LE$, $\forall x \in \Omega$ $w_{\neg\theta}(x) = f_\neg(w_\theta(x))$, $w_{\theta \wedge \varphi}(x) = f_\wedge(w_\theta(x), w_\varphi(x))$, $w_{\theta \vee \varphi}(x) = f_\vee(w_\theta(x), w_\varphi(x))$ and $w_{\theta \to \varphi}(x) = f_\to(w_\theta(x), w_\varphi(x))$ where $w_\theta(x)$ is shorthand for $w(\theta, x)$.

This should be contrasted with the condition of a calculus being weakly functional as defined below.

Definition 16. $w : LE \times \Omega \to [0,1]$ is said to be weakly functional iff $\forall \theta \in LE$ there exists a function $f_\theta : [0,1]^n \to [0,1]$ such that $w_\theta(x) = f_\theta(w_{L_1}(x), \ldots, w_{L_n}(x))$.

Clearly weak functionality is a strictly weaker condition than strong functionality. Strong functionality implies that the value of w for, say, a conjunction of expressions depends only on the value of w for the conjuncts and not on their logic structure. Weak functionality allows for that logical structure to be taken into account. It should be noted that weak functionality is sufficient to insure that all values for w can be determined from its values on LA and hence the amount of information needed to be stored is still of order n and not of order exponential in n or higher as is the case in many non-functional calculi (see [26]). In view of this we would argue that weak functionality is adequate to ensure computational feasibility for most real world applications.

In the literature, and especially in approximate reasoning it is often the case that the only type of functionality considered is strong functionality. However, clearly calculi exist that are weakly but not strongly functional; a typical example being a standard probabilistic calculus for which the basic events are assumed to be independent. This failure to distinguish clearly between these two levels of functionality can lead to misunderstandings. For example, consider the triviality results proved by Dubois and Prade [4] and later Elkan [9] which show that no non-binary functional calculus can satisfy all the laws of Boolean algebra. For example, only binary functional calculi can satisfy idempotence as well as the laws of excluded middle and non-contradiction (see [5]). It is important to realise that in this case the type of functionality referred to is strong functionality. Weakly functional calculi are not restricted in this manner; for instance probabilistic calculi under an independence assumption satisfy all the standard boolean laws while maintaining weak functionality.

To see that the appropriateness measure is not strongly functional notice that despite Propositions 13 and 14 it does not hold that $\forall\theta,\varphi \in LE$, $\mu^{\Gamma}_{\theta\wedge\varphi}(x) = \min(\mu^{\Gamma}_{\theta}(x), \mu^{\Gamma}_{\varphi}(x))$ or that $\mu^{\Gamma}_{\theta\vee\varphi}(x) = \max(\mu^{\Gamma}_{\theta}(x), \mu^{\Gamma}_{\varphi}(x))$. For instance, consider $\mu^{\Gamma}_{L_i\wedge\neg L_j}(x)$. From Definition 3 we have that $\lambda(L_i \wedge \neg L_j) = \lambda(L_i) \cap \overline{\lambda(L_j)}$ and hence

$$\mu^{\Gamma}_{L_i\wedge\neg L_j}(x) = \sum_{S:\ L_i\in S,\ L_j\notin S} m^{\Gamma}_x(S).$$

Given the consonance assumption we know that $m^{\Gamma}_x = M_0 : m_0, \ldots, M_k : m_k$ where $M_t \subset M_{t+1}$ for $t = 0, \ldots, k-1$. Now suppose that $\mu^{\Gamma}_{L_i}(x) \leqslant \mu^{\Gamma}_{L_j}(x)$ then for all $t = 0, \ldots, k$ if $L_i \in M_t$ then $L_j \in M_t$ and hence $\mu^{\Gamma}_{L_i\wedge\neg L_j}(x) = 0$. Alternatively if $\mu^{\Gamma}_{L_i}(x) \geqslant \mu^{\Gamma}_{L_j}(x)$ then

$$\begin{aligned}\mu^{\Gamma}_{L_i\wedge\neg L_j}(x) &= \sum_{S:\ L_i\in S,\ L_j\notin S} m^{\Gamma}_x(S) = \sum_{S:\ L_i\in S} m^{\Gamma}_x(S) - \sum_{S:\ L_j\in S} m^{\Gamma}_x(S)\\ &= \mu^{\Gamma}_{L_i}(x) - \mu^{\Gamma}_{L_j}(x).\end{aligned}$$

This can be summarised by the expression $\mu^{\Gamma}_{L_i\wedge\neg L_j}(x) = \max(0, \mu^{\Gamma}_{L_i}(x) - \mu^{\Gamma}_{L_j}(x))$ which is not in general the same as $\min(\mu^{\Gamma}_{L_i}(x), 1 - \mu^{\Gamma}_{L_j}(x))$ as would be given by the strongly functional calculus consistent with Propositions 12–14 for which $f_\wedge(a,b) = \min(a,b)$, $f_\vee(a,b) = \max(a,b)$ and $f_\neg(a) = 1-a$. As an aside, we note that this result gives some insight into the behaviour of implication in label semantics, at least at the level of individual labels. For instance, we have that $L_i \to L_j$ is logically equivalent to $\neg(L_i \wedge \neg L_j)$ and hence $\mu^{\Gamma}_{L_i\to L_j}(x) = 1 - \mu^{\Gamma}_{L_i\wedge\neg L_j}(x) = 1 - \max(0, \mu^{\Gamma}_{L_i}(x) - \mu^{\Gamma}_{L_j}(x)) = \min(1, 1 - \mu^{\Gamma}_{L_i}(x) + \mu^{\Gamma}_{L_j}(x))$. This corresponds to Lukasiewicz implication (see [14] or [18]) although it only applies here at the label level and not for more complex expressions.

$$\boxed{\{\mu_{L_1}(x),\ldots,\mu_{L_n}(x)\}} \Longrightarrow [m_x] \Longrightarrow \boxed{\mu_\theta(x) = f_\theta(\mu_{L_1}(x)\ldots\mu_{L_n}(x)) = \textstyle\sum_{S\in\lambda(\theta)} m_x(S)}$$

Fig. 1. Weak functionality of label semantics.

To see that appropriateness degrees are weakly functional recall from elementary random set theory that a consonant mass assignment [13] is uniquely defined by its fixed point coverage. This means that m_x^Γ can be completely determined from the values of $\mu_{L_1}^\Gamma(x),\ldots,\mu_{L_n}^\Gamma(x)$ as follows: Let $\{y_1,\ldots,y_k\} = \{\mu_L^\Gamma(x) \mid L \in LA,\ \mu_L^\Gamma(x) > 0\}$ ordered such that $y_t > y_{t+1}$ for $t = 1,\ldots,k-1$ then

$$m_x^\Gamma = M_t : y_t - y_{t+1}, \quad t = 1,\ldots k-1, \quad M_k : y_k, \quad M_0 : 1 - y_1,$$

where $M_0 = \emptyset$ and $M_t = \{L \in LA \mid \mu_L^\Gamma(x) \geqslant y_t\}$ for $t = 1,\ldots,k$. Hence, since from Definition 11 $\mu_\theta^\Gamma(x)$ is uniquely determined by $\lambda(\theta)$ and m_x^Γ then there is clearly a functional relationship between $\mu_\theta^\Gamma(x)$ and $\mu_{L_1}^\Gamma(x),\ldots,\mu_{L_n}^\Gamma(x)$ (see Fig. 1). In other words, for every linguistic expression θ there is a function $f_\theta : [0,1]^n \to [0,1]$ such that $\mu_\theta(x) = f_\theta(\mu_{L_1}(x)\ldots\mu_{L_n}(x))$ where f_θ is evaluated by using the consonance assumption to infer a mass assignment on label sets and then summing the masses of sets contained in $\lambda(\theta)$. It should also be noted that the appropriateness degree satisfies the laws of the excluded middle and non-contradiction in the sense that for any frame Γ, $\forall x \in \Omega$, $\forall \theta \in LE$ $\mu_{\theta\vee\neg\theta}^\Gamma(x) = 1$ and $\mu_{\theta\wedge\neg\theta}^\Gamma(x) = 0$ as follows immediately from Propositions 9 and 12. Thus the consonance assumption applied to label sets results in a functional calculus that coincides with the standard fuzzy logic connectives at the basic label level while preserving the laws of excluded middle and non-contradiction. This should be contrasted with the consonance assumption applied to random sets on the attribute universe which is not, in itself, sufficient to generate a functional calculus across a number of fuzzy concepts.

The weak functionality of label semantics brings considerable practical advantages since we no longer need to have any knowledge of the underlying population of individuals V or their distribution P_V (i.e., the frame) in order to determine m_x. Rather, for reasoning with label semantics in practice we need only define appropriateness degrees μ_L for $L \in LA$ corresponding to the imprecise definition of each label.

One possible method for calculating $\mu_\theta^\Gamma(x)$ for a general $\theta \in LE$ and $x \in \Omega$ is as follows: By the disjunctive normal form theorem we have that θ is logically equivalent to a disjunction of atoms $\bigvee_{\alpha:\,\alpha\to\theta}\alpha$ where each atom is a conjunction of literals of the form $\alpha = \bigwedge_i \pm L_i$. Now it can easily been seen, from Lemma 6, that for any atom of this form $\lambda(\alpha)$ is a singleton consisting of the subset of LA made up from those labels appearing positively in α. Also by Definition 3 and Corollary 8 we have that $\lambda(\theta) = \bigcup_{\alpha:\,\alpha\to\theta}\lambda(\alpha)$ and hence $\mu_\theta^\Gamma(x) = \sum_{\alpha:\,\alpha\to\theta} m_x^\Gamma(\lambda(\alpha))$. (NB. We are abusing notation slightly here and taking $\lambda(\alpha)$ to correspond to the single element of 2^{LA} associated with α rather than the set containing that element.) Alternatively, it may be more convenient just to determine $\lambda(\theta)$ recursively according to Definition 3.

In the specific context of a particular frame we may be able to make inferences regarding label expressions that do not generally hold. Furthermore, since a frame effectively defines the meaning and relationship between the members of LA, it identifies a subset of 2^{LA} as the sets of appropriate labels that can actually occur. This restriction will make the calculation of appropriateness degrees much less complex provided the basic labels do not overlap semantically too much. For instance, given $LA = \{small, medium, large\}$ we may find that in some frame Γ *small* only overlaps with *medium*, *medium* overlaps with *small* and *large* and *large* only overlaps with *medium*. This means that only the following occur as sets of possible labels: $\emptyset$, $\{small\}$, $\{small, medium\}$, $\{medium\}$, $\{medium, large\}$, $\{large\}$. We can formalise this observation by defining the set of focal elements for a frame as follows:

Definition 17 (*Set of focal elements*). The set of focal elements for frame Γ is $\mathcal{F}_\Gamma = \{S \subseteq LA \mid \exists x \in \Omega,\ m_x^\Gamma(S) > 0\}$.

In other words, the focal sets correspond to the sets of appropriate labels that are consistent with the definition of the labels in frame Γ. Given this concept we can define the following natural semantic relations on LE.

Definition 18.

(i) (Follows from in frame Γ) For $\theta, \varphi \in LE$ φ follows from θ in frame Γ (denoted $\theta \models_\Gamma \varphi$) iff $\lambda(\theta) \cap \mathcal{F}_\Gamma \subseteq \lambda(\varphi) \cap \mathcal{F}_\Gamma$.
(ii) (Equivalent to in frame Γ) For $\theta, \varphi \in LE$ φ is equivalent to θ in frame Γ (denoted $\varphi \equiv_\Gamma \theta$) iff $\lambda(\theta) \cap \mathcal{F}_\Gamma = \lambda(\varphi) \cap \mathcal{F}_\Gamma$.
(iii) For $\theta \in LE$ θ is universally true in frame Γ (denoted $\models_\Gamma \theta$) iff $\lambda(\theta) \cap \mathcal{F}_\Gamma = \mathcal{F}_\Gamma$.

The frame Γ provides an interpretation for each label in LA, as made apparent by their respective appropriateness degrees, and this should be taken into account in any subsequent reasoning. So, for instance, while it may not generally be the case that no value can be both $small \wedge large$, it certainly is true in any frame for which the appropriateness degrees of *small* and *large* do not overlap. The operators $\models_\Gamma$ and $\equiv_\Gamma$ incorporate this additional information on the meaning of labels into the logical notions of 'follows from' and 'equivalent to'. These operators can also be defined in terms of standard propositional logic deduction as the following result shows.

Definition 19.

$$\forall S \subseteq LA \quad \alpha_S = \left(\bigwedge_{L_i \in S} L_i\right) \wedge \left(\bigwedge_{L_i \notin S} \neg L_i\right).$$

Lemma 20. $\lambda(\bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S) = \mathcal{F}_\Gamma$.

Proof. $\lambda(\bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S) = \bigcup_{S \in \mathcal{F}_\Gamma} \lambda(\alpha_S)$ by Definition 3. Now

$$\begin{aligned}
\lambda(\alpha_S) &= \lambda\left(\left(\bigwedge_{L_i \in S} L_i\right) \wedge \left(\bigwedge_{L_i \notin S} \neg L_i\right)\right) \\
&= \bigcap_{L_i \in S} \lambda(L_i) \cap \bigcap_{L_i \notin S} \lambda(\neg L_i) \quad \text{by Definition 19} \\
&= \bigcap_{L_i \in S} \{T \subseteq LA \mid L_i \in T\} \cap \bigcap_{L_i \notin S} \{T \subseteq LA \mid L_i \notin T\} \quad \text{by Definition 3} \\
&= \{T \subseteq LA \mid S \subseteq T\} \cap \{T \subseteq LA \mid \overline{S} \cap T = \emptyset\} = \{S\}.
\end{aligned}$$

Therefore, $\lambda(\bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S) = \bigcup_{S \in \mathcal{F}_\Gamma} \{S\} = \mathcal{F}_\Gamma$ as required. □

Proposition 21. $\theta \models_\Gamma \varphi$ *iff* $(\bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S) \wedge \theta \models \varphi$.

Proof. $(\Rightarrow)$

$$\begin{aligned}
\theta \models_\Gamma \varphi &\Rightarrow \lambda(\theta) \cap \mathcal{F}_\Gamma \subseteq \lambda(\varphi) \cap \mathcal{F}_\Gamma \quad \text{by Definition 18} \\
&\Rightarrow \lambda(\theta) \cap \lambda\left(\bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) \subseteq \lambda(\varphi) \cap \lambda\left(\bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) \subseteq \lambda(\varphi) \quad \text{by Lemma 20} \\
&\Rightarrow \lambda\left(\theta \wedge \bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) \subseteq \lambda(\varphi) \quad \text{by Definition 3} \\
&\Rightarrow \left\{\tau(v) \mid v \in Val, v\left(\theta \wedge \bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) = true\right\} \\
&\Rightarrow \subseteq \{\tau(v) \mid v \in Val, v(\varphi) = true\} \quad \text{by Lemma 6} \\
&\Rightarrow \left\{v \in Val \mid v\left(\theta \wedge \bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) = true\right\} \\
&\Rightarrow \subseteq \{v \in Val \mid v(\varphi) = true\} \quad \text{since } \tau \text{ is a bijection} \\
&\Rightarrow \theta, \bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S \models \varphi \quad \text{as required.}
\end{aligned}$$

$(\Leftarrow)$

$$\theta, \bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S \models \varphi$$

$$\left\{v \in Val \mid v\left(\theta \wedge \bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) = true\right\} \subseteq \{v \in Val \mid v(\varphi) = true\}$$

$$\Rightarrow \left\{v \in Val \mid v\left(\theta \wedge \bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) = true\right\} \subseteq \left\{v \in Val \mid v\left(\varphi \wedge \bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) = true\right\}$$

$$\left\{\tau(v) \mid v \in Val, v\left(\theta \wedge \bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) = true\right\}$$

$$
\begin{aligned}
&\subseteq \left\{\tau(v) \mid v \in Val, v\left(\varphi \wedge \bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) = true\right\} \quad \text{since } \tau \text{ is a bijection}\\
&\Rightarrow \lambda\left(\theta \wedge \bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) \subseteq \lambda\left(\varphi \wedge \bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) \quad \text{by Lemma 6}\\
&\Rightarrow \lambda(\theta) \cap \lambda\left(\bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) \subseteq \lambda(\varphi) \cap \lambda\left(\bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S\right) \quad \text{by Definition 3}\\
&\Rightarrow \lambda(\theta) \cap \mathcal{F}_\Gamma \subseteq \lambda(\varphi) \cap \mathcal{F}_\Gamma \quad \text{by Lemma 20}\\
&\Rightarrow \theta \models_\Gamma \varphi \quad \text{as required.} \qquad \square
\end{aligned}
$$

Proposition 21 tells us that the information regarding the meaning of labels in LA contained in a particular focal set S can be completely represented by the logical expression α_S. In some sense this is to be expected since for any set S, α_S provides a logical description of S by stating exactly what labels are and are not contained in S. The next corollary follows trivially from Proposition 21.

Corollary 22. $\theta \equiv_\Gamma \varphi$ *iff*

$$\bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S \wedge \theta \equiv \bigvee_{S \in \mathcal{F}_\Gamma} \alpha_S \wedge \varphi.$$

We should observe that since m_x^Γ can be completely determined by $\mu_{L_1}^\Gamma(x), \ldots, \mu_{L_n}^\Gamma(x)$ then the set of focal elements for Γ can also be determined given only these values. Therefore, in a strong sense the full meaning of the labels in LA are captured by their appropriateness degrees. A common example of (i) in Definition 18 is when a certain label is conceptually implied by another label. For instance, we might say that whenever someone is described as being *very tall* then they can also be described as *tall*. In fuzzy set theory this would be captured by taking the fuzzy set for *very tall* as a fuzzy subset of the fuzzy set for *tall*. In label semantics we would expect to have a frame Γ in which whenever *very tall* was deemed an appropriate label so was *tall*. In other words *very tall* $\models_\Gamma$ *tall* or alternatively $\models_\Gamma$ *very tall* $\rightarrow$ *tall*. In such a case it is easy to see that $\forall x \in \Omega\ \mu_{\mathit{very\ tall}}^\Gamma(x) \leqslant \mu_{\mathit{tall}}^\Gamma(x)$ so that in this instance fuzzy set theory and label semantics would coincide. In general, we have that for any frame Γ such that $\theta \models_\Gamma \varphi$ then $\forall x \in \Omega$ $\mu_\theta^\Gamma(x) \leqslant \mu_\varphi^\Gamma(x)$.

Example 23. Let $LA = \{small, medium, large\}$, $\Omega = [0, 10]$ and Γ be a frame such that μ_{small}^Γ, μ_{medium}^Γ and μ_{large}^Γ are trapezoidal functions (see Fig. 2) defined by

$$\mu_{small}^\Gamma(x) = \begin{cases} 1 & \text{if } x \in [0,2],\\ 2 - \frac{x}{2} & \text{if } x \in (2,4],\\ 0 & \text{if } x > 4, \end{cases} \qquad \mu_{medium}^\Gamma(x) = \begin{cases} 0 & \text{if } x < 2,\\ \frac{x}{2} - 1 & \text{if } x \in (2,4],\\ 1 & \text{if } x \in (4,6],\\ 4 - \frac{x}{2} & \text{if } x \in (6,8],\\ 0 & \text{if } x > 8, \end{cases}$$

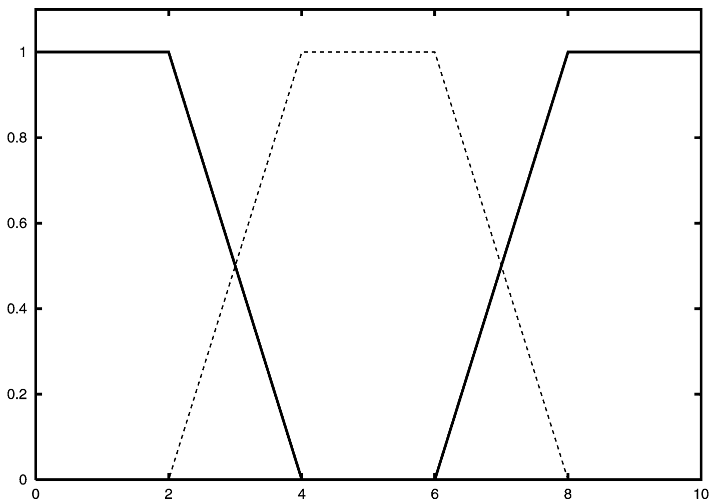

Fig. 2. Appropriateness degrees for, from left to right, *small*, *medium* and *large*.

$$\mu^{\Gamma}_{large}(x) = \begin{cases} 0 & \text{if } x < 6, \\ \frac{x}{2} - 3 & \text{if } x \in [6, 8], \\ 1 & \text{if } x > 8. \end{cases}$$

Allowing x to vary across $[0, 10]$ we obtain the following definition of m^{Γ}_x as follows: (see Fig. 3):

$$m^{\Gamma}_x\big(\{small\}\big) = \begin{cases} 1 & \text{if } x \in [0, 2], \\ 3 - x & \text{if } x \in (2, 3], \\ 0 & \text{if } x > 3, \end{cases}$$

$$m^{\Gamma}_x\big(\{small, medium\}\big) = \begin{cases} 0 & \text{if } x < 2, \\ \frac{x}{2} - 1 & \text{if } x \in [2, 3], \\ 2 - \frac{x}{2} & \text{if } x \in (3, 4], \\ 0 & \text{if } x > 4, \end{cases}$$

$$m^{\Gamma}_x\big(\{medium\}\big) = \begin{cases} 0 & \text{if } x < 3, \\ x - 3 & \text{if } x \in [3, 4], \\ 1 & \text{if } x \in (4, 6], \\ 7 - x & \text{if } x \in (6, 7], \\ 0 & \text{if } x > 7, \end{cases}$$

$$m^{\Gamma}_x\big(\{medium, large\}\big) = \begin{cases} 0 & \text{if } x < 6, \\ \frac{x}{2} - 3 & \text{if } x \in [6, 7], \\ 4 - \frac{x}{2} & \text{if } x \in (7, 8], \\ 0 & \text{if } x > 8, \end{cases}$$

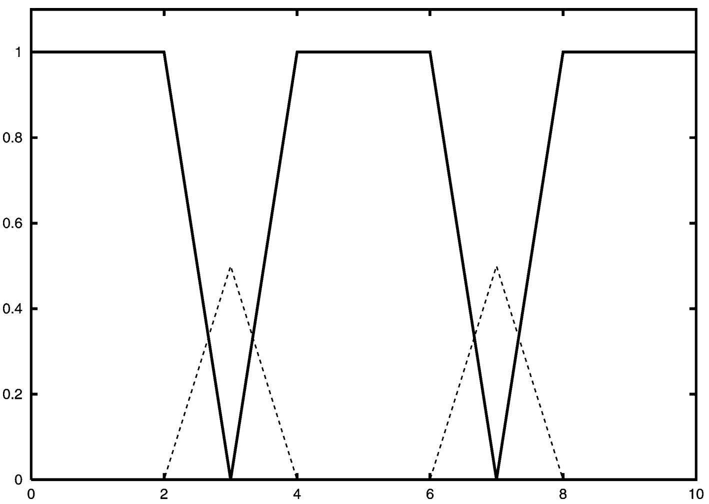

Fig. 3. Mass assignments for varying x; shown from left to right, $m_x^\Gamma(\{small\})$, $m_x^\Gamma(\{small, medium\})$, $m_x^\Gamma(\{medium\})$, $m_x^\Gamma(\{medium, large\})$ and $m_x^\Gamma(\{large\})$; $m_x^\Gamma(\emptyset)$ is equal to $m_x^\Gamma(\{small, medium\})$ for $x \in [2,4]$, is equal to $m_x^\Gamma(\{medium, large\})$ for $x \in [6,8]$ and is zero otherwise.

$$m_x^\Gamma\big(\{large\}\big) = \begin{cases} 0 & \text{if } x < 7, \\ x-7 & \text{if } x \in [7,8], \\ 1 & \text{if } x > 8, \end{cases} \qquad m_x^\Gamma(\emptyset) = \begin{cases} 0 & \text{if } x < 2, \\ \frac{x}{2}-1 & \text{if } x \in [2,3], \\ 2-\frac{x}{2} & \text{if } x \in (3,4], \\ 0 & \text{if } x \in (4,6], \\ \frac{x}{2}-3 & \text{if } x \in (6,7], \\ 4-\frac{x}{2} & \text{if } x \in (7,8], \\ 0 & \text{if } x > 8. \end{cases}$$

This gives a set of focal elements $\mathcal{F}_\Gamma = \{\emptyset, \{small\}, \{small, medium\}, \{medium\}, \{medium, large\}, \{large\}\}$ from which, for example, it follows that:

$$\mu^\Gamma_{medium \wedge \neg large}(x)$$
$$= m_x^\Gamma\big(\{small, medium\}\big) + m_x^\Gamma\big(\{medium\}\big) = \begin{cases} 0 & \text{if } x < 2, \\ \frac{x}{2}-1 & \text{if } x \in [2,4], \\ 1 & \text{if } x \in (4,6], \\ 7-x & \text{if } x \in (6,7], \\ 0 & \text{if } x > 7, \end{cases}$$

$$\mu^\Gamma_{\neg small \wedge medium \wedge \neg large}(x) = m_x^\Gamma\big(\{medium\}\big) = \begin{cases} 0 & \text{if } x < 3, \\ x-3 & \text{if } x \in [3,4], \\ 1 & \text{if } a \in (4,6], \\ 7-x & \text{if } x \in (6,7], \\ 0 & \text{if } x > 7, \end{cases}$$

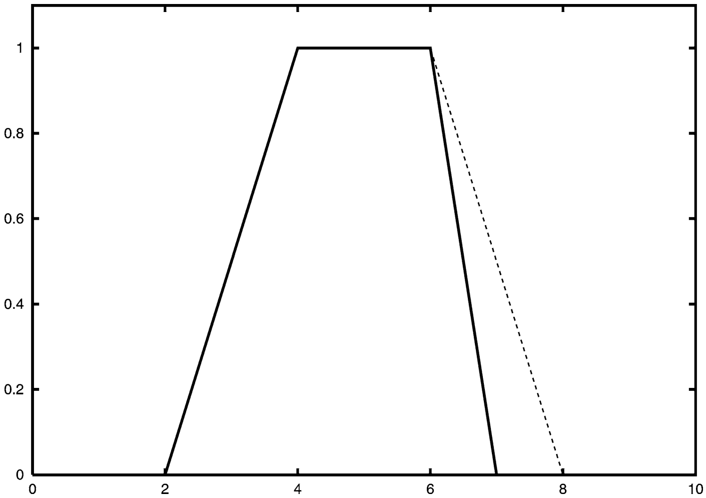

Fig. 4. Appropriateness degree $\mu^{\Gamma}_{medium\wedge\neg large}(x)$ (solid line) and $\min(\mu^{\Gamma}_{medium}(x), 1-\mu^{\Gamma}_{large}(x)) = \mu^{\Gamma}_{medium}(x)$ (dashed line).

$$\mu^{\Gamma}_{\neg(small\vee medium)}(x) = m^{\Gamma}_x\big(\{large\}\big) + m^{\Gamma}_x(\emptyset) = \begin{cases} 0 & \text{if } x < 2, \\ \frac{x}{2} - 1 & \text{if } x \in [2,3], \\ 2 - \frac{x}{2} & \text{if } x \in (3,4], \\ 0 & \text{if } x \in (4,6], \\ \frac{x}{2} - 3 & \text{if } x \in (6,8], \\ 1 & \text{if } x > 8. \end{cases}$$

Fig. 3 shows the values of the mass assignment m_x for each focal element as x ranges across Ω. From this we see that mass is associated with the empty set for values in the ranges $[2,4]$ and $[6,8]$. In label semantics this suggest that there are individuals in V for whom none of the terms in *LA* are appropriate as labels for values in these ranges. One might observe that this phenomena occurs frequently in natural language especially when labelling perceptions generated along some continuum. For example, we occasionally encounter colours for which none of our available colour descriptors seem appropriate. Fig. 4 clearly illustrates the difference between label semantics and fuzzy logic when evaluating compound expressions such as, in this case, $medium \wedge \neg large$. It is interesting to note that using strongly functional fuzzy logic based on min as the conjunction function, the two statements x *is medium* and x *is medium but not large* provide exactly the same information (i.e., they have the same memberships). In other words, the extra information that x *is not large* tells us nothing. This seems highly counter intuitive. On the other hand, in label semantics $\mu_{medium\wedge\neg large}$ is zero for all values greater than seven since for such values the only sets of appropriate labels with non-zero mass containing *medium* also contain *large*.

4. Multi-dimensional label semantics

Most modelling problems involve multiple attributes or variables. Therefore, if label semantics is to provide an effective knowledge representation framework for linguistic modelling it must be generalised to the multi-dimensional case. In other words, we need to provide a means of interpreting and evaluating linguistic expressions involving more than one variable.

Specifically, consider a modelling problem with k variables (or attributes) $x_1, \ldots, x_k$ with associated universes $\Omega_1, \ldots, \Omega_k$. For each variable we define a set of labels $LA_j = \{L_{1,j}, \ldots, L_{n_j,j}\}$ for $j = 1, \ldots, k$. In this case we ask individuals from V to provide a set of appropriate labels for each attribute value. Hence, an individual I will provide a vector of label descriptions $\langle \mathcal{D}^I_{a_1}, \ldots, \mathcal{D}^I_{a_k} \rangle$ for the attribute vector $\langle a_1, \ldots, a_k \rangle$. In this context we can extend the definitions of mass assignment and appropriateness degree given in Section 3 to the multi-dimensional case. Initially, however, we formally define k-dimensional linguistic expressions.

Let LE_j be the set of label expression for variable x_j generated by recursive application of the connectives $\wedge$, $\vee$, $\rightarrow$, $\neg$ to the labels in LA_j. We can now define the set of multi-dimensional label expression for describing linguistic relationships between variables as follows:

Definition 24 (*Multi-dimensional label expressions*). $MLE^{(k)}$ is the set of all multi-dimensional label expressions that can be generated from the label expression LE_j: $j = 1, \ldots, k$ and is defined recursively by:

(i) If $\theta \in LE_j$ for $j = 1, \ldots, k$ then $\theta \in MLE^{(k)}$.
(ii) If $\theta, \varphi \in MLE^{(k)}$ then $\neg\theta, \theta \wedge \varphi, \theta \vee \varphi, \theta \rightarrow \varphi \in MLE^{(k)}$.

Any k-dimensional label expression θ identifies a subset of $2^{LA_1} \times \cdots \times 2^{LA_k}$, denoted $\lambda^{(k)}(\theta)$, constraining the cross product of label descriptions $\mathcal{D}_{x_1} \times \cdots \times \mathcal{D}_{x_k}$. In this way the imprecise constraint θ on $x_1 \times \cdots \times x_k$ is interpreted as the precise constraint $\mathcal{D}_{x_1} \times \cdots \times \mathcal{D}_{x_k} \in \lambda^{(k)}(\theta)$.

Definition 25 (*Multi-dimensional appropriate label sets*). $\forall \theta \in MLE^{(k)}$ $\lambda^{(k)}(\theta) \subseteq 2^{LA_1} \times \cdots \times 2^{LA_k}$ such that

- $\forall \theta \in LE_j$ $\lambda^{(k)}(\theta) = {}^1\lambda(\theta) \times \bigtimes_{i \neq j} 2^{LA_i}$.
- $\forall \theta, \varphi \in MLE^{(k)}$ $\lambda^{(k)}(\theta \wedge \varphi) = \lambda^{(k)}(\theta) \cap \lambda^{(k)}(\varphi)$.
- $\lambda^{(k)}(\theta \vee \varphi) = \lambda^{(k)}(\theta) \cup \lambda^{(k)}(\varphi)$.
- $\lambda^{(k)}(\theta \rightarrow \varphi) = \overline{\lambda^{(k)}(\theta)} \cup \lambda^{(k)}(\varphi)$.
- $\lambda^{(k)}(\neg\theta) = \overline{\lambda^{(k)}(\theta)}$.

Note that in the context of a particular frame Γ it may be more convenient to evaluate $\lambda^{(k)}(\theta) \cap \bigtimes_{j=1}^{k} \mathcal{F}^{(j)}_{\Gamma}$ where $\mathcal{F}^{(j)}_{\Gamma}$ is the set of focal elements for LA_j given frame Γ (see Definition 17).

[1] $\lambda(\theta) \subseteq LA_j$ refers to the one dimensional appropriate label set as given in Definition 3.

Example 26. Consider a modelling problem with two variables x_1 and x_2 for which $LA_1 = \{small, medium, large\}$ and $LA_2 = \{low, moderate, high\}$. Also suppose that for a given frame Γ the focal elements for LA_1 and LA_2 are, respectively:

$$\mathcal{F}_\Gamma^{(1)} = \big\{\{small\},\ \{small, medium\},\ \{medium\},\ \{medium, large\},\ \{large\}\big\},$$
$$\mathcal{F}_\Gamma^{(2)} = \big\{\{low\},\ \{low, moderate\},\ \{moderate\},\ \{moderate, high\},\ \{high\}\big\}.$$

Now according to Definition 25 we have that:

$$\begin{aligned}\lambda^{(2)}\big((medium \wedge \neg small) \wedge \neg low\big) &= \lambda^{(2)}(medium \wedge \neg small) \cap \lambda^{(2)}(\neg low)\\ &= \lambda(medium \wedge \neg small) \times \lambda(\neg low).\end{aligned}$$

Now

$$\lambda(medium \wedge \neg small) \cap \mathcal{F}_\Gamma^{(1)} = \big\{\{medium\},\ \{medium, large\}\big\}$$

and

$$\lambda(\neg low) \cap \mathcal{F}_\Gamma^{(2)} = \big\{\{moderate\},\ \{moderate, high\},\ \{high\}\big\}.$$

Hence,

$$\begin{aligned}&\lambda^{(2)}\big((medium \wedge \neg small) \wedge \neg low\big) \cap \big(\mathcal{F}_\Gamma^{(1)} \times \mathcal{F}_\Gamma^{(2)}\big)\\ &\quad = \big\{\langle\{medium\},\ \{moderate\}\rangle,\\ &\qquad \langle\{medium\},\ \{moderate, high\}\rangle,\\ &\qquad \langle\{medium\},\ \{high\}\rangle,\\ &\qquad \langle\{medium, large\},\ \{moderate\}\rangle,\\ &\qquad \langle\{medium, large\},\ \{moderate, high\}\rangle,\\ &\qquad \langle\{medium, large\},\ \{high\}\rangle\big\}.\end{aligned}$$

Definition 27 (*Joint mass assignment*).

$$\forall x_j \in \Omega_j\ \forall S_j \subseteq LA_j\colon\ j = 1, \ldots, k \quad m_{\langle x_1,\ldots,x_k\rangle}(S_1, \ldots, S_k) = \prod_{j=1}^{k} m_{x_j}(S_j).$$

Now

$$m_{\langle x_1,\ldots,x_k\rangle}(S_1, \ldots, S_k) = P_V\big(\{I \in V\colon\ \mathcal{D}_{x_1} = S_1, \ldots, \mathcal{D}_{x_k} = S_k\}\big)$$

provided we make the following conditional independence assumption. It is assumed that for each individual I the choice of appropriate labels for variable x_j is dependent only on the value of x_j, once this is known, and is independent of the value of any other variables. This is actually quite a weak assumption and does not *a prior* imply independence between the variables.

Definition 28 (*Multi-dimensional appropriateness degrees*).

$$\forall\theta \in MLE^{(k)},\ \forall x_j \in \Omega_j\colon\ j=1,\dots,k$$
$$\mu_\theta^{(k)}(x_1,\dots,x_k) = \sum_{\langle S_1,\dots,S_k\rangle \in \lambda^{(k)}(\theta)} m_{\langle x_1,\dots,x_k\rangle}(S_1,\dots,S_k) = \sum_{\langle S_1,\dots,S_k\rangle \in \lambda^{(k)}(\theta)} \prod_{j=1}^{k} m_{x_j}(S_j).$$

Proposition 29. *If $\theta \in MLE^{(c)}$ for $c<k$ then*

$$\forall x_j \in \Omega_j\colon\ j=1,\dots,k \quad \mu_\theta^{(k)}(x_1,\dots,x_k) = \mu_\theta^{(c)}(x_1,\dots,x_c).$$

Proof. By Definition 25 $\lambda^{(k)}(\theta) = \lambda^{(c)}(\theta) \times \times_{j=c+1}^{k} 2^{LA_j}$ and therefore,

$$\begin{aligned}
\mu_\theta^{(k)}(x_1,\dots,x_k) &= \sum_{\langle S_1,\dots,S_c\rangle \in \lambda^{(c)}(\theta)} \sum_{\langle S_{c+1},\dots,S_k\rangle \in 2^{LA_{c+1}} \times\cdots\times 2^{LA_k}} \prod_{j=1}^{m} m_{x_j}(S_j)\\
&= \sum_{\langle S_1,\dots,S_c\rangle \in \lambda^{(c)}(\theta)} \prod_{j=1}^{c} m_{x_j}(S_j) \sum_{S_{c+1}\subseteq LA_{c+1}} \cdots \sum_{S_k \subseteq LA_k} \prod_{j=c+1}^{k} m_{x_j}(S_j)\\
&= \sum_{\langle S_1,\dots,S_c\rangle \in \lambda^{(c)}(\theta)} \prod_{j=1}^{c} m_{x_j}(S_j) \left[\prod_{j=c+1}^{k} \sum_{S_j \subseteq LA_j} m_{x_j}(S_j)\right]\\
&= \sum_{\langle S_1,\dots,S_c\rangle \in \lambda^{(c)}(\theta)} \prod_{j=1}^{c} m_{x_j}(S_j) = \mu_\theta^{(c)}(x_1,\dots,x_c)
\end{aligned}$$

as required. □

Proposition 30. *Let $\theta_j \in LE_j\colon\ j=1,\dots,k$, then the appropriateness degree of the conditional $(\bigwedge_{j=1}^{k-1}\theta_j) \to \theta_k$ is given by*

$$\mu^{(k)}_{(\bigwedge_{j=1}^{k-1}\theta_j)\to\theta_k}(x_1,\dots,x_k) = 1 - \prod_{j=1}^{k-1} \mu_{\theta_j}(x_j) + \prod_{j=1}^{k} \mu_{\theta_j}(x_j).$$

Proof. By Definition 25 we have that

$$\lambda^{(k)}\left(\left(\bigwedge_{j=1}^{k-1}\theta_j\right) \to \theta_k\right) = \overline{\lambda^{(k)}\left(\bigwedge_{j=1}^{k-1}\theta_j\right)} \cup \lambda^{(k)}(\theta_k) = \overline{\lambda^{(k)}\left(\bigwedge_{j=1}^{k-1}\theta_j\right) \cap \overline{\lambda^{(k)}(\theta_k)}}.$$

Now again by Definition 25 it follows that

$$\lambda^{(k)}\left(\bigwedge_{j=1}^{k-1}\theta_j\right) \cap \overline{\lambda^{(k)}(\theta_k)} = \lambda(\theta_1) \times \cdots \times \lambda(\theta_{k-1}) \times \overline{\lambda(\theta_k)}.$$

Therefore,

$$\begin{aligned}
&\mu^{(k)}_{(\bigwedge_{j=1}^{k-1}\theta_j)\to\theta_k}(x_1,\ldots,x_k)\\
&\quad = 1-\left[\prod_{j=1}^{k-1}\sum_{S_j\in\lambda(\theta_j)} m_{x_j}(S_j)\right]\times\left[\sum_{S_k\in\overline{\lambda(\theta_k)}} m_{x_k}(S_k)\right]\\
&\quad = 1-\left[\prod_{j=1}^{k-1}\mu_{\theta_j}(x_j)\right]\times\left[1-\mu_{\theta_k}(x_k)\right] = 1-\prod_{j=1}^{k-1}\mu_{\theta_j}(x_j)+\prod_{j=1}^{k}\mu_{\theta_j}(x_j)
\end{aligned}$$

as required. □

It is interesting to note that this corresponds to the use of a Reichenbach implication operator which, not surprisingly, is generated from the product *t-conorm*.

Example 31. Consider a modelling problem with two variables x_1, x_2 each with universe $[0, 10]$ and for which we have defined the label sets $LA_1 = \{small_1(s_1), medium_1(m_1), large_1(l_1)\}$ and $LA_2 = \{small_2(s_2), medium_2(m_2), large_2(l_2)\}$. For both variables the appropriateness degrees for *small*, *medium* and *large* are defined as in Example 23. Now suppose we learn that:

If x_1 is *medium* but not *large* then x_2 is *medium*

then according to Proposition 30 the appropriateness degree for $medium_1 \wedge \neg large_1 \to medium_2$ is given by

$$\begin{aligned}
&\mu^{(2)}_{medium_1\wedge\neg large_1\to medium_2}(x_1,x_2)\\
&\quad = 1-\mu_{medium_1\wedge\neg large_1}(x_1)+\mu_{medium_1\wedge\neg large_1}(x_1)\mu_{medium_2}(x_2).
\end{aligned}$$

Assuming the appropriateness degrees for *small*, *medium* and *large* given in Example 23 then the resulting function is shown in Fig. 5.

Clearly, this function provides information regarding the relationship between x_1 and x_2 assuming the constraint $medium_1 \wedge \neg large_1 \to medium_2$. For instance, from Fig. 5 we

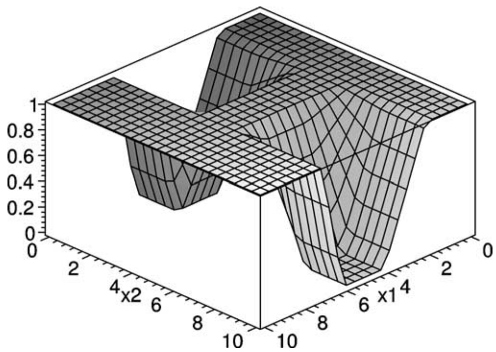

Fig. 5. Plot of the appropriateness degree for $medium_1 \wedge \neg large_1 \to medium_2$.

can see that if $x_1 = 5$ the it is very unlikely that $8 \leqslant x_2 \leqslant 10$. The problem of how to make output predictions (for x_2) given input values (for x_1) is considered in detail in the sequel.

Now suppose we also learn that

If x_1 is *large* then x_2 is *small*.

In this case we would want to evaluate the appropriateness degrees for the expression $(\mathit{medium}_1 \wedge \neg \mathit{large}_1 \rightarrow \mathit{medium}_2) \wedge (\mathit{large}_1 \rightarrow \mathit{small}_2)$. For this expression we have

$$\begin{aligned}
&\lambda^{(2)}\big((m_1 \wedge \neg l_1 \rightarrow m_2) \wedge (l_1 \rightarrow s_2)\big) \\
&\quad = \big(\overline{\lambda^{(2)}(m_1 \wedge \neg l_1)} \cup \lambda^{(2)}(m_2)\big) \cap \big(\overline{\lambda^{(2)}(l_1)} \cup \lambda^{(2)}(s_2)\big) \\
&\quad = \overline{\big(\lambda^{(2)}(m_1 \wedge \neg l_1) \cap \overline{\lambda^{(2)}(m_2)}\big) \cup \big(\lambda^{(2)}(l_1) \cap \overline{\lambda^{(2)}(s_2)}\big)} \\
&\quad = \overline{\big(\lambda(m_1 \wedge \neg l_1) \times \overline{\lambda(m_2)}\big) \cup \big(\lambda(l_1) \times \overline{\lambda(s_2)}\big)}.
\end{aligned}$$

Now

$$\begin{aligned}
&\lambda(m_1 \wedge \neg l_1) \times \overline{\lambda(m_2)} \cap \big(\mathcal{F}_\Gamma^{(1)} \times \mathcal{F}_\Gamma^{(2)}\big) \\
&\quad = \big\{\{s_1, m_1\},\ \{m_1\}\big\} \times \big\{\{m_2\},\ \{m_2, l_2\},\ \{l_2\},\ \emptyset\big\}
\end{aligned}$$

and

$$\lambda(l_1) \times \overline{\lambda(s_2)} \cap \big(\mathcal{F}_\Gamma^{(1)} \times \mathcal{F}_\Gamma^{(2)}\big) = \big\{\{l_1\},\ \{m_1, l_1\}\big\} \times \big\{\{m_2\},\ \{m_2, l_2\},\ \{l_2\},\ \emptyset\big\}.$$

Hence, since these two sets on $2^{LA_1} \times 2^{LA_2}$ are mutually exclusive it follows that:

$$\begin{aligned}
&\forall x_1 \in \Omega_1,\ \forall x_2 \in \Omega_2\ \mu^{(2)}_{(m_1 \wedge \neg l_1 \rightarrow m_2) \wedge (l_1 \rightarrow s_2)}(x_1, x_2) \\
&\quad = 1 - \Bigg(\sum_{S_1 \in \lambda(m_1 \wedge \neg l_1)} \sum_{S_2 \in \overline{\lambda(m_2)}} m_{x_1}(S_1) m_{x_2}(S_2) + \sum_{S_1 \in \lambda(l_1)} \sum_{S_2 \in \overline{\lambda(s_2)}} m_{x_1}(S_1) m_{x_2}(S_2)\Bigg) \\
&\quad = 1 - \big(\mu_{m_1 \wedge \neg l_1}(x_1) \times \big(1 - \mu_{m_2}(x_2)\big) + \mu_{l_1}(x_1) \times \big(1 - \mu_{s_2}(x_2)\big)\big).
\end{aligned}$$

Again assuming the appropriateness degrees for *small*, *medium* and *large* given in Example 23 then the resulting function is shown in Fig. 6.

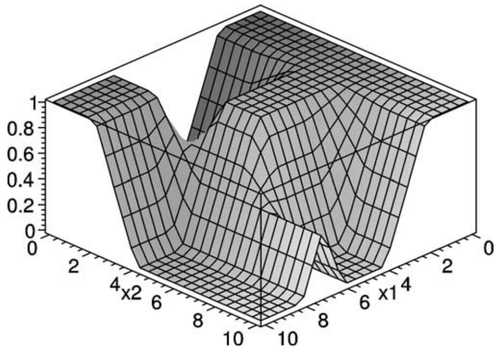

Fig. 6. Plot of the appropriateness degree for $(\mathit{medium}_1 \wedge \neg \mathit{large}_1 \rightarrow \mathit{medium}_2) \wedge (\mathit{large}_1 \rightarrow \mathit{small}_2)$.

The semantics proposed in this section are based on the idea that the meaning of vague linguistic expressions are determined by their use across a population of individuals. This is very close to the theory of vagueness proposed by Black [3]. An alternative viewpoint is that fuzzy concepts are inherently vague independent of their actual use. In principle, it may be possible to provide an operational semantics for membership functions consistent with this interpretation but very little foundational work of this kind has been undertaken. For example, one possible semantics of this kind is based on the idea that membership values are a measure of similarity to some set of prototypical exemplars of the concept (see [31,33]). However, such an alternative approach to the problem of vague concepts is not within the scope of this paper where instead we focus entirely on the random set interpretation.

5. Conditional information from linguistic constraints

To understand what is the information content of linguistic expressions we must also consider the nature of the constraints that such expression place on the underlying variable. If it is know that *Bill is tall* exactly what does this tell us about Bill's height? For example, can we determine an exact value, a distribution of values or a family of distributions? In [37] it is proposed that such constraints specify a possibility distribution on the underlying variable, namely that given by the membership degree of the associated fuzzy set. This in itself would suggest a resulting family of probability distributions as characterised by the corresponding possibility and dual necessity measure. However, in many applications of fuzzy sets, in particular fuzzy control, so-called defuzzification techniques seem to treat the possibility distribution as if it were a probability distribution in order to estimate a precise value for the variable. In the case of Bill's height the so called centre of mass defuzzification method (see [30]) would evaluate

$$\frac{\int_{\Omega} x \mu_{tall}(x)\,\mathrm{d}x}{\int_{\Omega} \mu_{tall}(x)\,\mathrm{d}x}.$$

Clearly, this has no obvious semantic justification in fuzzy set theory. In addition, the association of membership values with fuzzy sets as discussed in [37] is more in the nature of a primitive definition rather than being a consequence of some lower level semantics for either membership or possibility.

In this section we will introduce a label semantics based approach to linguistic constraints for which we will argue that in order to make any inferences about the underlying variable based on a linguistic expression we must not only have knowledge of the frame but also of the associated extended frame. Since appropriateness degrees, the analogue of fuzzy memberships in label semantics, are determined by the frame only then clearly we are claiming that such information alone is inadequate to draw any but the most general conclusions from linguistic expressions. For simplicity, we will assume in the sequel that the extended frame Γ^+ is such that either P_Ω is discrete or it has an associated density function p_Ω. Now consider a knowledge base consisting of a single label expression θ with meaning $\mathcal{D}_x \in \lambda(\theta)$. Then according to Bayes' theorem we can infer the following posterior distribution on Ω:

– *Continuous case*.

$$\forall a \in \Omega \quad p(a|\theta) = \frac{Pr(\theta \mid x = a)\, p_\Omega(a)}{\int_\Omega Pr(\theta \mid x)\, p_\Omega(x)\, \mathrm{d}x}.$$

– *Discrete case*.

$$\forall a \in \Omega \quad Pr(x = a|\theta) = \frac{Pr(\theta \mid x = a)\, P_\Omega(x = a)}{\sum_{x \in \Omega:\ P_\Omega(x) > 0} Pr(\theta \mid x)\, P_\Omega(x)}.$$

Now according to label semantics

$$Pr(\theta \mid x = a) = Pr\big(\mathcal{D}_x \in \lambda(\theta) \mid x = a\big) = Pr\big(\mathcal{D}_a \in \lambda(\theta)\big) = \sum_{S \in \lambda(\theta)} m_a(S) = \mu_\theta(a).$$

Therefore, we obtain in the continuous case

$$\forall x \in \Omega \quad p(x \mid \theta) = \frac{\mu_\theta(x)\, p_\Omega(x)}{\int_\Omega \mu_\theta(x)\, p_\Omega(x)\, \mathrm{d}x}$$

and in the discrete case

$$\forall x \in \Omega \quad Pr(x \mid \theta) = \frac{\mu_\theta(x)\, P_\Omega(x)}{\sum_{x \in \Omega:\ P_\Omega(x) > 0} \mu_\theta(x)\, P_\Omega(x)}.$$

From the above we can see that appropriateness degrees may be viewed as a likelihood measure (i.e., $\mu_\theta(x)$ can be interpreted as the likelihood that θ is an appropriate label for x). This is not surprising since as Dubois and Prade [6] comment, the likelihood and random set semantics for fuzzy concepts are strongly linked. A number of authors have independently investigated likelihood semantics for possibility or fuzzy sets outside the random set framework, including Hisdal [17] and Dubois, Moral and Prade [7]. Also Thomas [32] has investigated the relationship between Bayesian reasoning and fuzzy sets.

The above likelihood interpretation has important consequences regarding the level of condition information we can obtain from the knowledge that x is constrained by θ. Clearly given that appropriateness degrees are defined for *LA* (i.e., we have sufficient knowledge of the frame) then our knowledge of the above posterior distribution depends entirely on our knowledge of the prior P_Ω (i.e., the associated extended frame). For example, if we know only that $P_\Omega \in \mathcal{P}$, for some set of distributions $\mathcal{P}$, then we will only be able to determine upper and lower bounds for the posterior describing an inferred family of posterior distributions. In the discrete case these upper and lower probabilities are defined as follows:

$$\forall x \in \Omega \quad Pr^*(x \mid \theta) = \sup_{P_\Omega \in \mathcal{P}} \frac{\mu_\theta(x)\, P_\Omega(x)}{\sum_{x \in \Omega} \mu_\theta(x)\, P_\Omega(x)}$$

and

$$\forall x \in \Omega \quad Pr_*(x \mid \theta) = \inf_{P_\Omega \in \mathcal{P}} \frac{\mu_\theta(x)\, P_\Omega(x)}{\sum_{x \in \Omega} \mu_\theta(x)\, P_\Omega(x)}.$$

This essentially, corresponds to a special case of imprecise Bayesian inference as proposed by Walley [34]. However, it should be noted that often in practice surprisingly little information can be inferred from such knowledge. For example, consider the scenario described in Example 1, but where our prior knowledge is that the score on the dice is either 2 or 3 (i.e., $\mathcal{P} = \{P_\Omega \mid P_\Omega(2) + P_\Omega(3) = 1\}$). Furthermore, suppose that we are also informed that the *SCORE is low*. Now since the appropriateness degree of *low* for 2 is 1 while for 3 it is only $\frac{1}{3}$ one might expect that we could safely infer that

$$Pr(SCORE = 2 \mid low) \geqslant Pr(SCORE = 3 \mid low).$$

A little trivial mathematics, however, reveals that this is not the case since for a prior where $P_\Omega(2) = 1$ we find that $Pr(SCORE = 2 \mid low) = 1$ and $Pr(SCORE = 3 \mid low) = 0$, while for a prior where $P_\Omega(3) = 1$ we obtain $Pr(SCORE = 3 \mid low) = 1$ and $Pr(SCORE = 2 \mid low) = 0$. Obviously, then in terms of upper and lower bounds we can infer only that $Pr(SCORE = 2 \mid low), Pr(SCORE = 3 \mid low) \in [0, 1]$ and $Pr(SCORE \in \{1, 2\} \mid low) = 1$. We see then that even in the presence of relatively specific linguistic constraints the information we can infer about the value of the underlying variable is strongly dependent on our prior assumptions about its distribution. For instance, the inequality

$$Pr(SCORE = 2 \mid low) \geqslant Pr(SCORE = 3 \mid low)$$

as suggested above, holds if and only if

$$P_\Omega(2) \geqslant \frac{\mu_{low}(3)}{\mu_{low}(2) + \mu_{low}(3)} = \frac{1}{4}.$$

See [35] for an alternative semantics for linguistic concepts based on upper and lower probabilities.

A common, although sometimes problematic (see [36] for a discussion) assumption in Bayesian analysis is to assume a uniform prior on Ω. In this case we obtain

$$Pr(2 \mid low) = \frac{\mu_{low}(2)}{\mu_{low}(2) + \mu_{low}(3)} = \frac{3}{4}$$

and

$$Pr(3 \mid low) = \frac{\mu_{low}(3)}{\mu_{low}(2) + \mu_{low}(3)} = \frac{1}{4}.$$

Generally, the assumption of a uniform prior on Ω gives us $Pr(x|\theta)$ proportional to $\mu_\theta(x)$, that is

$$p(x \mid \theta) = \frac{\mu_\theta(x)}{\int_\Omega \mu_\theta(x)\,\mathrm{d}x}$$

in the continuous case and

$$Pr(x \mid \theta) = \frac{\mu_\theta(x)}{\sum_{x \in \Omega} \mu_\theta(x)}$$

in the discrete case.

Now if it is required that we estimate a precise value of x on the basis of a linguistic constraint θ then a natural approach in the current context would be simply to determine the expected value of $Pr(x \mid \theta)$. Clearly, if a uniform prior on Ω is assumed then this gives us an expression for appropriateness degrees equivalent to the centre of mass defuzzification method as described above.

We now illustrate how such conditioning can be used to evaluate output values of a system given specific input values when the relationships between input and output variables are described in terms of linguistic expressions. Initially, however, we observe that the conditional distributions defined above can easily be extended to the multi-dimensional case so that for $\theta \in MLE^{(k)}$:

$$\forall x_j \in \Omega_j,\ j = 1, \ldots, k,$$
$$p(x_1, \ldots, x_k \mid \theta) = \frac{\mu_\theta^{(k)}(x_1, \ldots, x_k)\, p(x_1, \ldots, x_k)}{\int_{\Omega_1} \cdots \int_{\Omega_k} \mu_\theta^{(k)}(x_1, \ldots, x_k)\, p(x_1, \ldots, x_k)\, \mathrm{d}\vec{x}},$$

where $p(x_1, \ldots, x_k)$ is the prior distribution on $\Omega_1 \times \cdots \times \Omega_k$.

Example 32. Recall the problem described in Example 31 where our knowledge of the relationship between variables x_1 and x_2 corresponds to

$$\mathcal{K} \equiv (medium_1 \wedge \neg large_1 \rightarrow medium_2) \wedge (large_1 \rightarrow small_2).$$

Assuming a uniform prior distribution on $[0, 10]^2$ then the posterior distribution on $x_1 \times x_2$ is given by

$$\forall x_1 \in \Omega_1,\ x_2 \in \Omega_2$$
$$p(x_1, x_2 \mid \mathcal{K}) = \frac{\mu^{(2)}_{(m_1 \wedge \neg l_1 \rightarrow m_2) \wedge (l_1 \rightarrow s_2)}(x_1, x_2)}{\int_0^{10} \int_0^{10} \mu^{(2)}_{(m_1 \wedge \neg l_1 \rightarrow m_2) \wedge (l_1 \rightarrow s_2)}(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2}.$$

Now suppose we are given the value of x_1 and we want to calculate the probability of different values of x_2 given this information. In this case we need to evaluate the following conditional distribution:

$$p(x_2 \mid \mathcal{K}, x_1) = \frac{p(x_1, x_2 \mid \mathcal{K})}{p(x_1 \mid \mathcal{K})} = \frac{\mu^{(2)}_{(m_1 \wedge \neg l_1 \rightarrow m_2) \wedge (l_1 \rightarrow s_2)}(x_1, x_2)}{\int_0^{10} \mu^{(2)}_{(m_1 \wedge \neg l_1 \rightarrow m_2) \wedge (l_1 \rightarrow s_2)}(x_1, x_2)\, \mathrm{d}x_2}.$$

A plot of this distribution as both x_1 and x_2 vary is given in Fig. 7.

In the case that $x_1 = 6.5$ the conditional density $p(x_2 \mid \mathcal{K}, 6.5)$ is shown in Fig. 8.

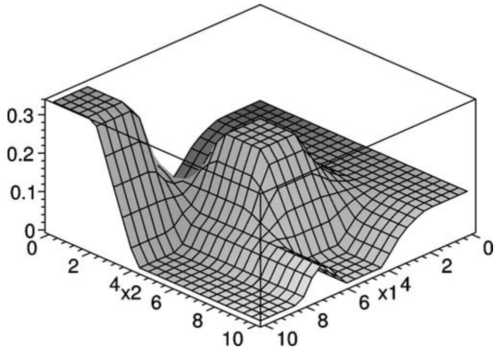

Fig. 7. Plot of the conditional density $p(x_2 \mid \mathcal{K}, x_1)$.

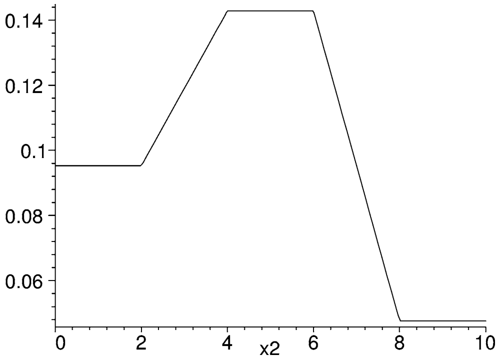

Fig. 8. Plot of the conditional density $p(x_2 \mid \mathcal{K}, 6.5)$.

Therefore, in order to obtain an estimate of output x_2 given input $x_1 = 6.5$ we can evaluate the expected value of this distribution:

$$\hat{x}_1 = \int_0^{10} x_2 p(x_2 \mid \mathcal{K}, 6.5)\, \mathrm{d}x_2 = 4.5079.$$

In many situations our conditional information may not take the form of a linguistic expression but rather a distribution on linguistic expressions. In label semantics such information provides constraints on the distribution (mass assignment) for $\mathcal{D}_x$ as x varies. Here we consider only the simplest case where sufficient constraints are available to specify a unique mass assignment on $\mathcal{D}_x$. To illustrate how such specific knowledge might be obtained let us return to the height problem where we have an extended frame $\Gamma^+ = \langle V, P_V, \Omega, P_\Omega \rangle$ where P_Ω is a prior based on a known distribution on heights of European males. Furthermore, suppose we have a database *DB* of heights of a finite number of British males so that for $x \in \Omega$, $P_{DB}(x)$ corresponds to the probability of a male of height x being chosen at random from *DB*. Given this we can evaluate a mass assignment on $\mathcal{D}_x$ conditional on the information that x is the height of a British male as follows:

$$\forall S \subseteq LA \quad m_{DB}(S) = \sum_{x \in \Omega:\ P_{DB}(x) > 0} P_{DB}(x) m_x(S).$$

Now given a posterior mass assignment m_{DB} on $\mathcal{D}_x$ what information can we infer regarding the underlying variable x? According to the theorem of total probability

$$p(a) = \sum_{S \subseteq LA} Pr(\mathcal{D}_x = S) p(a \mid \mathcal{D}_x = S).$$

Hence, if we know that $\forall S \subseteq LA\ Pr(\mathcal{D}_x = S) = m_{DB}(S)$ then we can condition on this knowledge as follows:

$$p(a \mid m_{DB}) = \sum_{S \subseteq LA} m_{DB}(S) p(a \mid \mathcal{D}_x = S).$$

Now according to Bayes theorem

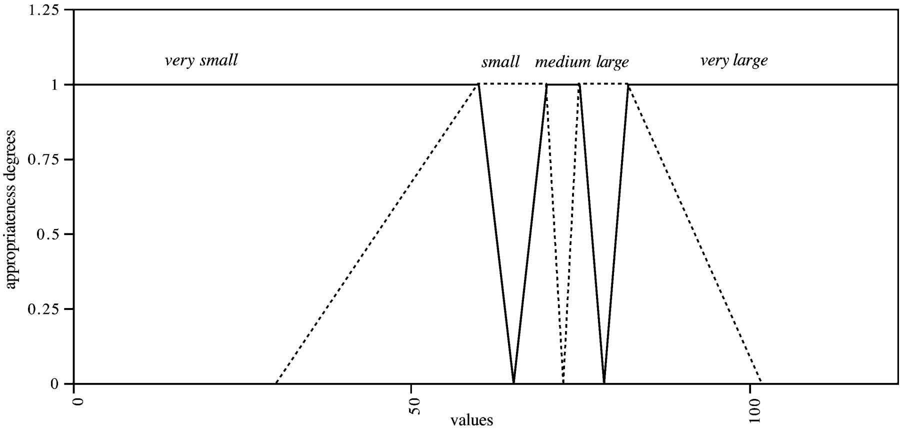

Fig. 9. Appropriateness degrees for labels of diastolic blood pressure.

$$\forall a \in \Omega \quad p(a \mid \mathcal{D}_x = S) = \frac{Pr(\mathcal{D}_x = S \mid x = a)\, p_\Omega(a)}{\int_\Omega Pr(\mathcal{D}_x = S \mid x = a)\, p_\Omega(a)\, \mathrm{d}a} = \frac{m_a(S)\, p_\Omega(a)}{\int_\Omega m_a(S)\, p_\Omega(a)\, \mathrm{d}a}.$$

Let

$$\int_\Omega m_a(S)\, p_\Omega(a)\, \mathrm{d}a = pm(S)$$

be the prior mass assignment on $\mathcal{D}_x$ generated by prior distribution P_Ω on Ω then we have

$$\forall x \in \Omega \quad p(x \mid m_{DB}) = \sum_{S \subseteq LA} m_{DB}(S) \frac{m_x(S)\, p_\Omega(x)}{pm(S)} = p_\Omega(x) \sum_{S \subseteq LA} \frac{m_{DB}(S)}{pm(S)} m_x(S).$$

Notice that in the case where $\forall S \subseteq LA\ m_{DB}(S) = pm(S)$, in other words when our posterior knowledge of $\mathcal{D}_x$ matches our prior knowledge, then $p(x \mid m_{DB}) = p_\Omega(x)$ as one would intuitively expect.

Example 33. This example relates to a database stored as part of the machine learning repository at the University of California at Irvine. It is essentially a classification problem but serves well to illustrate the use of label semantics to determine underlying distributions from data. The database itself contains the details of 768 females from the population of Pima Indians living near Phoenix Arizona, USA. The diagnostic binary-valued variable investigated is whether the patient shows signs of diabetes according to World Health Organisation criteria. There are eight measured variables which include, number of times pregnant, plasma glucose concentration, diastolic blood pressure, triceps skin fold thickness, 2-hour serum insulin, body mass index, diabetes pedigree function and age. A label semantics approach has been used in [21] and [29] in conjunction with a Bayesian classifier but here we shall simply use this example to show how a posterior mass assignment can be used to infer a posterior density on the underlying variable,

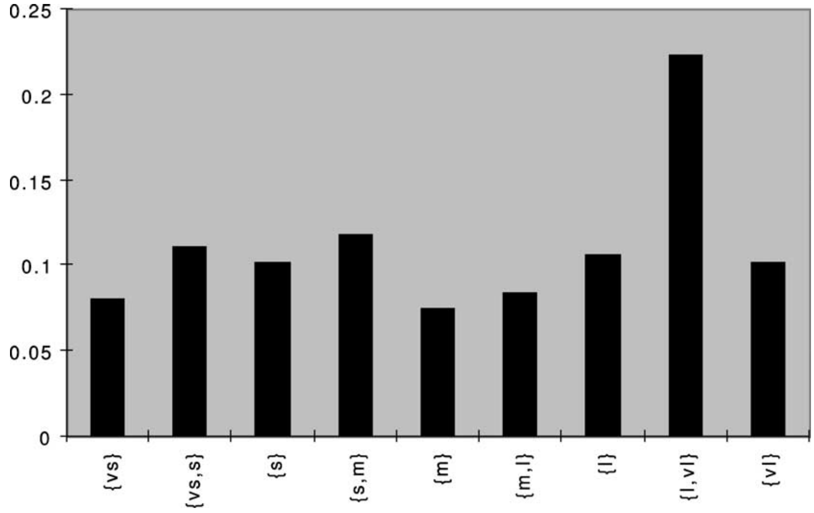

Fig. 10. Mass assignment on labels for the diastolic blood pressure of diabetics.

diastolic blood pressure. In this case $LA = \{very\ small, small, medium, large, very\ large\}$ with appropriateness degrees as shown in Fig. 9. These functions have been defined using a percentile based approach to ensure that each label covers approximately the same number of data elements. Clearly, the set of focal elements for this frame is given by

$$\mathcal{F}_\Gamma = \big\{\{very\ small\},\ \{small, very\ small\},\ \{small\},\ \{small, medium\},\ \{medium\},\ \{medium, large\},\ \{large\},\ \{large, very\ large\},\ \{very\ large\}\big\}.$$

The extended frame is assumed to be such that $\Omega = [0, 122]$ and P_Ω is the uniform distribution on this interval. The posterior mass assignments (see Fig. 10) generated from the sub-database, *DIAB*, of diabetic individuals is given by

$$\begin{aligned} m_{DIAB} = {} & \{very\ small\} : 0.0805969,\ \{very\ small, small\} : 0.110448,\\ & \{small\} : 0.101492,\ \{small, medium\} : 0.117911,\ \{medium\} : 0.07462,\\ & \{medium, large\} : 0.084223,\ \{large\} : 0.106608,\\ & \{large, very\ large\} : 0.222976,\ \{very\ large\} : 0.101119. \end{aligned}$$

The prior mass assignment for this domain based on a uniform prior P_Ω is

$$\begin{aligned} pm = {} & \{very\ small\} : 0.368,\ \{very\ small, small\} : 0.1434,\ \{small\} : 0.04099,\\ & \{small, medium\} : 0.03074,\ \{medium\} : 0.02049,\ \{medium, large\} : 0.02459,\\ & \{large\} : 0.02869,\ \{large, very\ large\} : 0.09631,\ \{very\ large\} : 0.2459. \end{aligned}$$

Now, for instance, if $a = 68$ we have that

$$m_{68} = \{small\} : 0.4,\ \{small, medium\} : 0.6$$

and therefore,

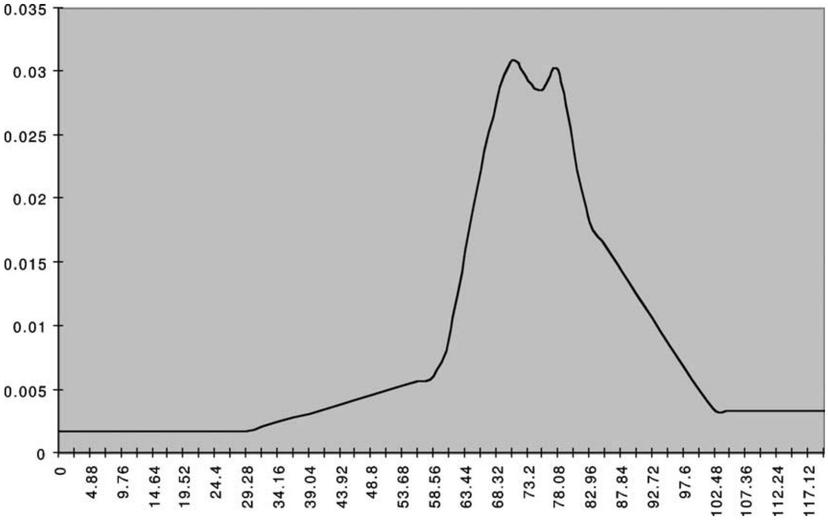

Fig. 11. Posterior density on diastolic blood pressure for diabetics.

$$
\begin{aligned}
p(68 \mid m_{DIAB}) &= \frac{1}{122}\left[\frac{m_{DIAB}(\{small\})}{pm(\{small\})} m_{68}\big(\{small\}\big)\right. \\
&\quad \left. + \frac{m_{DIAB}(\{small, medium\})}{pm(\{small, medium\})} m_{68}\big(\{small, medium\}\big)\right] \\
&= \frac{1}{122}\left[\frac{0.101492}{0.04099}(0.4) + \frac{0.11791}{0.03074}(0.6)\right] = 0.0269822.
\end{aligned}
$$

The full posterior density obtained is shown in Fig. 11.

Another interesting issue relating to conditional inference with linguistic expressions is that of the conditional matching of expressions. For example, suppose we are told that *Bill is tall* then with what level of certainty, if any, can we infer that *Bill is very tall*? In the next section we propose two approaches to matching within the framework of label semantics, both probabilistic in nature.

6. Matching linguistic expressions

Suppose it is known that the variable x is constrained by the linguistic expression φ. In this case, what is the degree to which another expression, θ, can appropriately be used to describe x. This is an important question that takes on special significance in the area of fuzzy or possibilistic logic programming [1,2,8]. In this context a mechanism is required by which we can evaluate the semantic match (or unification) of an expression θ, forming part of a query, with a given expression φ in the knowledge base. A number of authors have investigate this problem but of most relevance to the current framework is work by Baldwin et al. [2] who introduces a measure of semantic unification based on the conditional probability of fuzzy events. This measure is also based on random sets, but defined on the attribute universe rather than at the label level. In this section we present two measures of matching between expressions and discuss their respective properties. The first approach is as follows: If we know that linguistic expression φ holds then this corresponds to the event $\mathcal{D}_x \in \lambda(\varphi)$ and, according to Bayesian inference, we should update our prior mass assignment m to obtain a posterior mass assignment m_φ as follows:

$$\forall S \subseteq LA \quad m_\varphi(S) = \begin{cases} \dfrac{pm(S)}{\sum_{S\in\lambda(\varphi)} pm(S)} & \text{if } S \in \lambda(\varphi), \\ 0 & \text{otherwise.} \end{cases}$$

Interestingly m_φ can be used to show the consistency between the definition of conditional distribution given a linguistic expression and that of conditional distribution given a mass assignment as is highlighted by the following proposition. (NB. In the following two proofs it is assumed P_Ω has a density p_Ω. The finite case can be proved in a similar way.)

Proposition 34. $\forall x \in \Omega \;\; p(x|m_\varphi) = p(x \mid \varphi)$.

Proof.

$$\forall x \in \Omega \quad p(x \mid m_\varphi) = p_\Omega(x) \sum_{S\subseteq LA} \frac{m_\varphi(S)}{pm(S)} m_x(S) = p_\Omega(x) \sum_{S\in\lambda(\varphi)} \frac{m_\varphi(S)}{pm(S)} m_x(S)$$

by the definition of m_φ

$$= p_\Omega(x) \sum_{S\in\lambda(\varphi)} \frac{m_x(S)}{\sum_{S\in\lambda(\varphi)} pm(S)} = p_\Omega(x) \frac{\sum_{S\in\lambda(\varphi)} m_x(S)}{\sum_{S\in\lambda(\varphi)} pm(S)}.$$

Now $\sum_{S\in\lambda(\varphi)} m_x(S) = \mu_\varphi(x)$ by Definition 11 and

$$\sum_{S\in\lambda(\varphi)} pm(S) = \sum_{S\in\lambda(\varphi)} \int_\Omega m_x(S) p_\Omega(x)\,\mathrm{d}x = \int_\Omega \Bigg(\sum_{S\in\lambda(\varphi)} m_x(S)\Bigg) p_\Omega(x)\,\mathrm{d}x$$
$$= \int_\Omega \mu_\varphi(x) p_\Omega(x)\,\mathrm{d}x.$$

Therefore

$$p(x \mid m_\varphi) = \frac{p_\Omega(x)\mu_\varphi(x)}{\int_\Omega p_\Omega(x)\mu_\varphi(x)\,\mathrm{d}x} = p(x \mid \varphi). \qquad \square$$

Given m_φ we can then evaluate the likelihood of another linguistic expression θ according to the following definition.

Definition 35 (*Matching of type I*).

$$\forall \theta, \varphi \in LE \quad \mu^{\Gamma+}_{\theta|\varphi} = \sum_{S\in\lambda(\theta)} m_\varphi(S).$$

It can easily be seen that this definition of match can be expressed in terms of conditional probabilities on $\mathcal{D}_x$ as follows:

Proposition 36. $\forall\theta,\varphi \in LE$ $\mu^{\Gamma+}_{\theta|\varphi} = Pr(\mathcal{D}_x \in \lambda(\theta) \mid \mathcal{D}_x \in \lambda(\varphi))$.

Proof.

$$Pr\big(\mathcal{D}_x \in \lambda(\theta) \mid \mathcal{D}_x \in \lambda(\varphi)\big) = \frac{Pr(\mathcal{D}_x \in \lambda(\theta) \cap \lambda(\varphi))}{Pr(\mathcal{D}_x \in \lambda(\varphi))} = \frac{Pr(\mathcal{D}_x \in \lambda(\theta \wedge \varphi))}{Pr(\mathcal{D}_x \in \lambda(\varphi))}$$

by Definition 3 and by Bayes' theorem

$$= \frac{\int_\Omega Pr(\mathcal{D}_x \in \lambda(\theta \wedge \varphi) \mid x = a)\, p_\Omega(a)\,\mathrm{d}a}{\int_\Omega Pr(\mathcal{D}_x \in \lambda(\varphi) \mid x = a)\, p_\Omega(a)\,\mathrm{d}a} = \frac{\int_\Omega (\sum_{S\in\lambda(\theta\wedge\varphi)} m_a(S))\, p_\Omega(a)\,\mathrm{d}a}{\int_\Omega (\sum_{S\in\lambda(\varphi)} m_a(S))\, p_\Omega(a)\,\mathrm{d}a}$$
$$= \frac{\sum_{S\in\lambda(\theta\wedge\varphi)} \int_\Omega m_a(S)\, p_\Omega(a)\,\mathrm{d}a}{\sum_{S\in\lambda(\varphi)} \int_\Omega m_a(S)\, p_\Omega(a)\,\mathrm{d}a} = \frac{\sum_{S\in\lambda(\theta\wedge\varphi)} pm(S)}{\sum_{S\in\lambda(\varphi)} pm(S)}$$
$$= \sum_{S\in\lambda(\theta\wedge\varphi)} \frac{pm(S)}{\sum_{S\in\lambda(\varphi)} pm(S)} = \sum_{S\in\lambda(\theta)} m_\varphi(S). \qquad \square$$

Given the above it can easily be seen that

$$\mu^{\Gamma+}_{\theta|\varphi} = \frac{\int_\Omega \mu_{\theta\wedge\varphi}(x)\, p_\Omega(x)\,\mathrm{d}x}{\int_\Omega \mu_\varphi(x)\, p_\Omega(x)\,\mathrm{d}x}.$$

Interestingly, when $\theta,\varphi \in LA$ this corresponds to the degree of subsethood measure as proposed by Kosko [19], although for compound expressions this is not the case. Also, this proposition shows that $\mu^{\Gamma+}_{\theta|\varphi}$ truly is the conditional extension of the general appropriateness measure defined in Section 3, as the notation suggests. Furthermore, notice that trivially $\mu^{\Gamma+}_{\varphi|\varphi} = 1$ and $\mu^{\Gamma+}_{\neg\varphi|\varphi} = 0$ while this is not the case for many definitions of conditional match proposed in the literature (for example, [2]). However, in label semantics there is an alternative definition for match that does not satisfy these properties, defined as follows. Suppose we know that φ has been asserted by some individual in V then what is the likelihood that θ will hold true for some other individual randomly chosen from V. To evaluate this we observe that given φ we can determine a distribution on the underlying variable x, $p(x \mid \varphi)$, and also that for any value of x we know the probability that θ will be deemed an appropriate label expression, $\mu_\theta(x)$. From this we obtain the following definition for the match of θ given φ.

Definition 37 (*Matching of type II*).

$$\forall\theta,\varphi \in LE \quad \pi^{\Gamma+}_{\theta|\varphi} = \int_\Omega \mu_\theta(x)\, p(x \mid \varphi)\,\mathrm{d}x.$$

Given the definition of $p(x|\varphi)$ this can be rewritten as

$$\pi^{\Gamma+}_{\theta|\varphi} = \frac{\int_\Omega \mu_\theta(x)\mu_\varphi(x)\, p_\Omega(x)\,\mathrm{d}x}{\int_\Omega \mu_\varphi(x)\, p_\Omega(x)\,\mathrm{d}a}.$$

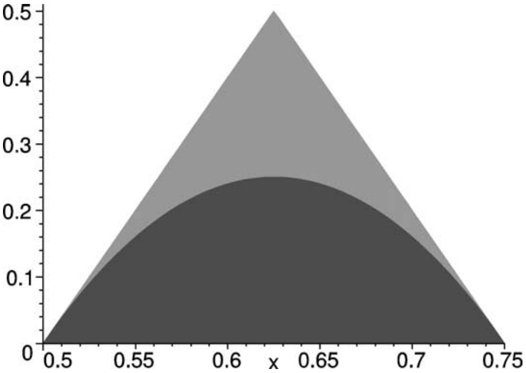

Fig. 12. Matching of types I and II for *large* given *medium*: The dark grey area corresponds to the numerator of $\pi^{\Gamma+}_{large|medium}$ assuming a uniform prior and the sum of the dark and light grey areas corresponds to the numerator of $\mu^{\Gamma+}_{large|medium}$.

Interestingly, taking μ to be analogous to fuzzy memberships then this corresponds to the definition of conditional probability of fuzzy events proposed in [37]. Obviously, in this case $\pi^{\Gamma+}_{\varphi|\varphi}$ is not necessarily equal to 1 (or indeed $\pi^{\Gamma+}_{\neg\varphi|\varphi}$ to 0). However, this is quite intuitive given the prevailing interpretation for π since the fact that a particular individual deems φ to be an appropriate label expression does not guarantee that all individuals will.

Example 38. Consider an extended frame for which $\Omega = [0,1]$, P_Ω is the uniform distribution on $[0,1]$ and

$$\mu_{medium} = \begin{cases} \frac{x-0.25}{0.25} & x \in [0.25, 0.5), \\ \frac{0.75-x}{0.25} & x \in [0.5, 0.75], \\ 0 & \text{otherwise}, \end{cases} \qquad \mu_{large} = \begin{cases} \frac{x-0.5}{0.25} & x \in [0.5, 0.75), \\ \frac{1-x}{0.25} & x \in [0.75, 1], \\ 0 & \text{otherwise}, \end{cases}$$

then we have that

$$\mu^{\Gamma+}_{large|medium} = \frac{\int_0^1 \mu_{large\wedge medium}(x)\,\mathrm{d}x}{\int_0^1 \mu_{medium}(x)\,\mathrm{d}x} = \frac{\int_0^1 \min(\mu_{medium}(x), \mu_{large}(x))\,\mathrm{d}x}{\int_0^1 \mu_{medium}(x)\,\mathrm{d}x}$$
$$= \frac{0.0625}{0.25} = 0.25.$$

Alternatively,

$$\pi^{\Gamma+}_{large|medium} = \frac{\int_0^1 \mu_{large}(x)\mu_{medium}(x)\,\mathrm{d}x}{\int_0^1 \mu_{medium}(x)\,\mathrm{d}x} = \frac{\int_{0.5}^{0.75} (\frac{0.75-x}{0.25})(\frac{x-0.5}{0.25})\,\mathrm{d}x}{0.25}$$
$$= \frac{0.0416667}{0.25} = 0.166667.$$

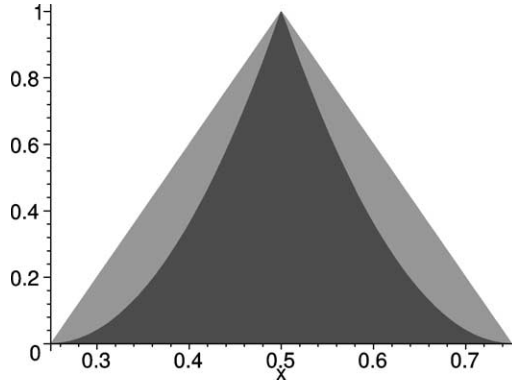

Fig. 13. Matching of type II for *medium* given *medium*: The dark grey area represents the numerator in $\pi^{\Gamma+}_{medium|medium}$.

Now as stated above $\mu^{\Gamma+}_{medium|medium} = 1$ but on the other hand

$$\pi^{\Gamma+}_{medium|medium} = \frac{\int_0^1 \mu^2_{medium}(x)\,\mathrm{d}x}{\int_0^1 \mu_{medium}(x)\,\mathrm{d}x} = \frac{\int_{0.25}^{0.5} (\frac{x-0.25}{0.25})^2\,\mathrm{d}x}{0.25} = \frac{0.166667}{0.25} = 0.666667.$$

7. Conclusions

A new framework for linguistic modelling, referred to as label semantics, has been presented, based on a random set interpretation of the measure of appropriateness of a label for a value. A natural, weakly functional calculus for appropriateness degree, has been described, which satisfies the law of the excluded middle and in general takes account of the logical structure of compound expressions when evaluating them. This calculus can then be combined with a bayesian framework to provide a means of inferring distributions on the underlying variable given both linguistic expressions and mass assignments. Furthermore, given a linguistic expression we have also presented methods for evaluating the likely applicability of other linguistic expressions based on the measures of conditional match types I and II.

Overall we would claim that label semantics has the potential to act as an effective high level knowledge representation framework for many modelling problems. At present applications have centred on its use in data mining and machine learning where a number of new methods have been developed based on the ideas proposed in Section 4 (see Lawry [21] and Randon [29]). In this context label semantics offers the prospect of combining both numerical and linguistic reasoning as discussed in Lawry [23]. Furthermore, it provides a mechanism for conditioning on prior or background linguistic information to infer probability distributions which can then be used in conjunction with models derived from data. In a different context, the method for estimating distributions from data described in Section 4 has also been used to evaluate imprecise probabilities of failure for risk analysis in environmental engineering (see [15] and [16]). More generally, the framework outlined in this paper gives us a coherent calculus for linguistic reasoning that may be used in a variety of decision-support problems. Certainly, in this context, the notions of appropriateness degree and conditional match will have a central role to play.

Acknowledgements

Many thanks to Tru Cao and John Shepherdson for their helpful discussions. I would also like to thank one of the anonymous referees for their patience and insightful comments. This work is partially funded by a grant from the Nuffield Foundation.

References

[1] T. Alsinet, L. Godo, A complete calculus for possibilistic logic programming with fuzzy propositional variable, in: Proceedings of Uncertainty in AI 2000, Stanford, CA, 2000.
[2] J.F. Baldwin, T.P. Martin, B.W. Pilsworth, Fril—Fuzzy and Evidential Reasoning in AI, Wiley, New York, 1995.
[3] M. Black, Vagueness: An exercise in logical analysis, Philos. Sci. 4 (1937) 427–455.
[4] D. Dubois, H. Prade, An introduction to possibility and fuzzy logics, in: P. Smets, et al. (Eds.), Non-Standard Logics for Automated Reasoning, Academic Press, New York, 1988, pp. 742–755.
[5] D. Dubois, H. Prade, Can we enforce full compositionality in uncertainty calculi?, in: Proc. AAAI-94, Seattle, WA, 1994, pp. 149–154.
[6] D. Dubois, H. Prade, The three semantics of fuzzy sets, Fuzzy Sets and Systems 90 (1997) 141–150.
[7] D. Dubois, S. Moral, H. Prade, A semantics for possibility theory based on likelihoods, J. Math. Anal. Appl. 205 (1997) 359–380.
[8] D. Dubois, H. Prade, Possibility theory: Qualitative and quantitative aspects, in: D.M. Gabbay, P. Smets (Eds.), Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, Kluwer, Dordrecht, 1998, pp. 169–226.
[9] C. Elkan, The paradoxical success of fuzzy logic, in: Proc. AAAI-93, Washington, DC, MIT Press, Cambridge, MA, 1993, pp. 698–703.
[10] B.R. Gaines, Fuzzy and probability uncertainty logics, J. Inform. Control 38 (1978) 154–169.
[11] H. Geffner, Default Reasoning: Causal and Conditional Theories, MIT Press, Cambridge, MA, 1992.
[12] I.R. Goodman, Fuzzy sets as equivalence classes of random sets, in: R. Rager (Ed.), Fuzzy Set and Possibility Theory, 1982, pp. 327–342.
[13] I.R. Goodman, H.T. Nguyen, Uncertainty Models for Knowledge Based Systems, North-Holland, Amsterdam, 1985.
[14] P. Hajek, Fuzzy logic from the logical point of view, in: M. Bartosek, et al. (Eds.), SOFSEM 95: Theory and Practice of Informatics, in: Lecture Notes in Computer Science, vol. 1012, 1995, pp. 31–49.
[15] J. Hall, J. Lawry, Imprecise probabilities of engineering system failure from random and fuzzy set reliability analysis, in: Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications, New York, 2001.
[16] J. Hall, J. Lawry, Fuzzy label methods for constructing imprecise limit state functions, Structural Safety 28 (2003) 317–341.
[17] E. Hisdal, Are grades of membership probabilities, Fuzzy Sets and Systems 25 (1988) 325–348.
[18] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic, Prentice-Hall, Englewood Cliffs, NJ, 1995.
[19] B. Kosko, Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence, Prentice-Hall, Englewood Cliffs, NJ, 1992.
[20] J. Lawry, A voting mechanism for fuzzy logic, Internat. J. Approx. Reason. 19 (1998) 315–333.

[21] J. Lawry, Label prototypes for modelling with words, in: Proceedings of The North American Fuzzy Information Processing Society 2001 Conference, 2001.
[22] J. Lawry, Label semantics: A formal framework for modelling with words, in: S. Benferhat, P. Besnard (Eds.), Proceedings of Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, in: Lecture Notes in Artificial Intelligence, vol. 2143, Springer, Berlin, 2001, pp. 374–384.
[23] J. Lawry, Query evaluation from linguistic prototypes, in: Proceedings of the FUZZ-IEEE 2001 Workshop on Modelling with Words, Melbourne, Australia, 2001, pp. 39–42.
[24] J.W. Lloyd, Foundations of Logic Programming, Second Edition, Springer, Berlin, 1987.
[25] H.T. Nguyen, On modeling of linguistic information using random sets, Inform. Sci. 34 (1984) 265–274.
[26] J.B. Paris, The Uncertain Reasoners Companion: A Mathematical Perspective, Cambridge University Press, Cambridge, 1994.
[27] J.B. Paris, Semantics for fuzzy logic supporting truth functionality, in: V. Novak, I. Perfilieva (Eds.), Discovering the World with Fuzzy Logic, Springer, Berlin, 2000.
[28] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Mateo, CA, 1988.
[29] N.J. Randon, J. Lawry, A transparent framework for data mining using modelling with words, in: Proceedings of 2001 UK Workshop on Computational Intelligence, 2001.
[30] E.H. Ruspini, P.P. Bonnisone, W. Pedtycz (Eds.), Handbook of Fuzzy Computation, Institute of Physics Publishing, 1998.
[31] E.H. Ruspini, On the semantics of fuzzy logic, Internat. J. Approx. Reason. 5 (1991) 45–88.
[32] S.F. Thomas, Fuzziness and Probability, ACG Press, Kansas, 1995.
[33] E. Trillas, L. Valverde, An enquiry into indistinguishability operators, in: H.J. Skala, S. Termini, E. Trillas (Eds.), Aspects of Vagueness, Kluwer Academic Publishers, Dordrecht, 1984, pp. 231–256.
[34] P. Walley, Statistical Inference with Imprecise Probabilities, Chapman and Hall, London, 1991.
[35] P. Walley, G. de Cooman, A behavioural model of linguistic uncertainty, Inform. Sci. 34 (1999) 1–37.
[36] S.L. Zabell, Symmetry and its discontents, in: Causation, Chance, and Credence, Vol. 1, Kluwer Academic, Dordrecht, 1988, pp. 155–190.
[37] L.A. Zadeh, Probability measures of fuzzy events, J. Math. Anal. Appl. 23 (1968) 421–427.
[38] L.A. Zadeh, The concept of linguistic variable and its application to approximate reasoning, Part 1, Inform. Sci. 8 (1975) 199–249.
[39] L.A. Zadeh, The concept of linguistic variable and its application to approximate reasoning, Part 2, Inform. Sci. 8 (1975) 301–357.
[40] L.A. Zadeh, The concept of linguistic variable and its application to approximate reasoning, Part 3, Inform. Sci. 9 (1976) 43–80.
[41] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1978) 3–28.
[42] L.A. Zadeh, Fuzzy logic = computing with words, IEEE Trans. Fuzzy Systems 2 (1996) 103–111.