Artificial Intelligence 173 (2009) 1539–1558

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Uncertainty modelling for vague concepts: A prototype theory approach

Jonathan Lawry [a,*], Yongchuan Tang [b]

[a] *Department of Engineering Mathematics, University of Bristol, Bristol BS8 1TR, UK*
[b] *College of Computer Science, Zhejiang University, Hangzhou 310027, PR China*

ARTICLE INFO

Article history:
Received 26 April 2008
Received in revised form 28 July 2009
Accepted 29 July 2009
Available online 5 August 2009

Keywords:
Epistemic vagueness
Prototype theory
Label semantics
Random sets

ABSTRACT

An epistemic model of the uncertainty associated with vague concepts is introduced. Label semantics theory is proposed as a framework for quantifying an agent's uncertainty concerning what labels are appropriate to describe a given example. An interpretation of label semantics is then proposed which incorporates prototype theory by introducing uncertain thresholds on the distance between elements and prototypes for description labels. This interpretation naturally generates a functional calculus for appropriateness measures. A more general model with distinct threshold variables for different labels is discussed and we show how different kinds of semantic dependence can be captured in this model.

1. Introduction

Natural language is a powerful, flexible and robust mechanism for communicating ideas, concepts and information. Yet the meaning conveyed by even simple words is often inherently uncertain. This uncertainty is reflected in the variation and inconsistency in the use of words by different individuals. For example, Parikh [30] reports an experiment where a sample of people are shown a chart with different coloured squares and asked to count the number of red and the number of blue squares. The results differ significantly across the group. Similar inconsistencies in the use of colour categories are also described in the work of Belin and Kay [1] and Kintz et al. [20]. We believe that this uncertainty about the appropriate use of words arises as a natural consequence of the distributed and case-based manner by which an understanding of language is acquired.

Language is, to a large degree, learnt through the experience of our interactions with other speakers from which we can make inferences about the implicit rules and conventions of language use [29]. Exposure to formal grammar rules and explicit dictionary definitions comes relatively late in our education and requires *a priori* a basic vocabulary on the part of the student. It is perhaps not surprising then that such a process results in significant semantic uncertainty. We cannot realistically expect that the boundaries of linguistic concepts, as perhaps represented by their extensions in a multi-dimensional conceptual space [11], should be precisely and unambiguously defined by a finite set of often conflicting examples. It is our view then, that the uncertainty about word meanings which naturally result from such an empirical learning process is the underlying source of concept vagueness. Consequently we adopt an epistemic perspective on vagueness, to some extent in accordance with the views of Williamson [37], whereby crisp concept boundaries are assumed to exist but where their precise definition is uncertain. Furthermore, as pointed out by Parikh [29,30], empirical learning requires extrapolation from previously encountered examples of word use to other new but similar cases. Hence, the notion of *similarity* is also fundamental to any model of vagueness. Prototype theory [32,33] provides a powerful tool to understand the role of *typicality* in concept definitions, resulting in a natural ordering on possible exemplars of concepts e.g. Bill is *taller* than Mary, but

* Corresponding author.
E-mail addresses: j.lawry@bris.ac.uk (J. Lawry), tyongchuan@gmail.com (Y. Tang).

0004-3702/$ – see front matter
doi:10.1016/j.artint.2009.07.006

Mary is *richer* than Bill. In this paper we attempt to provide a formal framework for representing the epistemic uncertainty associated with vague concepts, which incorporates elements of prototype theory.

The modelling of concept vagueness in Artificial Intelligence has been dominated by ideas from fuzzy set theory as originally proposed by Zadeh [38]. In that approach the extension of a concept is represented by a fuzzy set which has a graded characteristic or *membership function* with values ranging between 0 and 1. This allows for intermediate membership (values in $(0, 1)$) in vague concepts resulting in intermediate truth values for propositions involving vague concepts (fuzzy logic). The calculus for fuzzy set theory is truth-functional which means that the full complement of Boolean laws cannot all be satisfied [4].[1] Furthermore, fuzzy set theory and fuzzy logic do not in their narrowest manifestations adopt an epistemic view of vagueness. Hájek [14], for example, argues that membership values of fuzzy categories are primitives quantifying gradedness of membership according to which it is meaningless to refer to unknown or uncertain crisp boundaries of vague concepts, since such boundaries are inherently fuzzy. On the other hand, many of the proposed interpretations of fuzzy sets implicitly adopt an epistemic position. In particular, the random set model of fuzzy sets (see [12,13] and [27]) according to which fuzzy set membership functions correspond to single point coverage functions of a random set, inherently assumes the existence of an uncertain but crisp set representing the extension of a vague concept. The basis of the *label semantics* theory [21] outlined in this paper is also a random set model of vagueness but where the intention is to quantify uncertainty concerning the applicability or appropriateness of labels to describe a given example. Such a theory cannot result in a truth-functional calculus but can be functional in a weaker sense in the presence of certain assumptions concerning the semantic dependence between labels.

A principal motivation for this paper is to explore the relationship between prototype theory and label semantics. Hence, there will be a focus on mathematical results demonstrating a clear link between these two theories in the case when categorization (labeling) involves thresholding of a measure of similarity to prototypes. Furthermore, we will argue from an Artificial Intelligence perspective, that the proposed framework could be a suitable model for rational intelligent agents who use concept labels and label expressions to describe elements of their environment with the aim of communicating information to their fellow agents.

An outline of the paper is as follows: Section 2 describes a variant of the epistemic theory of vagueness which provides the philosophical underpinnings for the formal models we propose [23]. Section 3 provides an overview of label semantics as first proposed in [21] and [22]. Section 4 discusses the relationship between prototype theory, typicality, uncertainty and vagueness. Section 5 describes a new prototype theory interpretation of label semantics and finally Section 6 gives some conclusions and possible directions for future work.

2. An epistemic theory of vagueness

In our everyday use of language we are continually faced with decisions about the best way to describe objects and instances in order to convey the information we intend. For example, suppose you are witness to a robbery, how should you describe the robber so that police on patrol in the streets will have the best chance of spotting him? You will have certain labels that can be applied, for example *tall, short, medium, fat, thin, blonde, etc.*, some of which you may view as inappropriate for the robber, others perhaps you think are definitely appropriate while for some labels you are uncertain whether they are appropriate or not. On the other hand, perhaps you have some ordered preferences between labels so that *tall* is more appropriate than *medium* which is in turn more appropriate than *short*. Your choice of words to describe the robber should surely then be based on these judgments about the appropriateness of labels. Yet where does this knowledge come from and more fundamentally what does it actually mean to say that a label is or is not appropriate? In the sequel we shall propose an interpretation of vague description labels based on a particular notion of appropriateness and suggest a measure of subjective uncertainty resulting from an agent's partial knowledge about what labels are appropriate to assert. Furthermore, we will suggest that the vagueness of these description labels lies fundamentally in the uncertainty about if and when they are appropriate as governed by the rules and conventions of language use.

It seems undeniable that humans posses some kind of mechanism for deciding whether or not to make certain assertions (e.g. 'The robber is blonde') or to agree to a classification (e.g. 'Yes he was tall'). Furthermore, although the underlying concepts are often vague the decisions about assertions are, at a certain level, bivalent. That is to say for a particular example x and description θ, you are either willing to assert that 'x is θ' or not. Of course in general this decision may depend on many factors associated with the context in which the communication is taking place. For example, you are likely to be much more cautious in your use of language when describing a robber to the police than in describing a colleague to a close friend. Also, your motives may be much more complex than purely to communicate information. For example, you may have recognized the robber as a family member so that your aim when describing him is to throw the police off the scent. Nonetheless, there seems to be an underlying assumption that some things can be correctly asserted while others cannot. Exactly where the dividing line lies between those labels that are and those that are not appropriate to use may be uncertain, but the assumption that such a division exists would be a natural precursor to any decision making process of the kind just described.

[1] Except in the case where truth-values are restricted to $\{0, 1\}$ when fuzzy set theory reduces to classical set theory.

The above argument brings us very close to the epistemic view of vagueness as expounded by Timothy Williamson [37]. Williamson assumes that for the extensions of a vague concept there is a precise but unknown dividing boundary between it and the extension of the negation of that concept. For example, consider the set of heights which are classified as being tall, then there is according to the epistemic view a precise but unknown height for which all values less than this height are not tall while all those greater than it are tall. From this viewpoint Sorities problems are resolved by denying the assumption that practically indistinguishable elements satisfy the same vague predicates. Hence, for a finite sequence of increasing heights x_i: $i = 1, \ldots, k$ where x_1 is not tall, x_k is tall and $x_{i+1} - x_i \leqslant \epsilon$, for some very small positive number ϵ, it holds that: $\exists i$ for which 'x_i is not tall' and 'x_{i+1} is tall'. Although the exact value of i will be virtually impossible for anyone to identify precisely.

2.1. The epistemic stance

While there are marked similarities between the epistemic view and that proposed in this paper, there are also some important differences. For instance, the epistemic view would seem to assume the existence of some objectively correct, but unknown, set of criteria for determining whether or not a given instance satisfies a vague concept. Instead of this we argue that individuals when faced with decision problems regarding assertions find it useful as part of a decision making strategy to assume that there is a clear dividing line between those labels which are and those which are not appropriate to describe a given instance. In other words, in deciding what to assert agents behave *as if* the epistemic view of vagueness is correct. We refer to this strategic assumption across a population of communicating agents as the *epistemic stance* [22,23], a concise statement of which is as follows:

> *Each individual agent in the population assumes the existence of a set of labeling conventions, valid across the whole population, governing what linguistic labels and expressions can be appropriately used to describe particular instances.*

In practice these rules and conventions underlying the appropriate use of labels would not be imposed by some outside authority. In fact, they may not exist at all in a formal sense. Rather they are represented as a distributed body of knowledge concerning the assertability of predicates in various cases, shared across a population of agents, and emerging as the result of interactions and communications between individual agents all adopting the epistemic stance. The idea is that the learning processes of individual agents, all sharing the fundamental aim of understanding how words can be appropriately used to communicate information, will eventually converge to some degree on a set of shared conventions. The very process of convergence then to some extent vindicates the epistemic stance from the perspective of individual agents. Of course, this is not to suggest complete or even extensive agreement between individuals as to these appropriateness conventions. However, the overlap between agents should be sufficient to ensure the effective transfer of useful information. Indeed, such effective communication does not require perfect agreement. For example, [30] illustrates how two individuals with different notions of the colour blue can still effectively use the concept to pass information between them, by considering how the search space of one person looking for a book required by the other is reduced when they learn that the book is blue.

In many respects our view is quite close to that of Rohit Parikh [29,30] where he argues for an anti-representational view of vagueness, focusing on the notion of assertability rather than that of truth. Parikh argues that it is almost unavoidable that different speakers will use the same predicate in different ways because of the manner in which language is learnt. Since vague predicates lack a clear definition we tend to learn the 'usage of these words in some few cases and then we extrapolate'. With reference to Dewey and Wittgenstein, Parikh argues for a view of language where truth is relatively unimportant, but where communication is best thought of in terms of a set of social practices. What is important then is not whether a particular expression is true but whether it is assertible. To quote Parikh directly:

> "Certain sentences are assertible in the sense that we might ourselves assert them and other cases of sentences which are non-assertible in the sense that we ourselves (and many others) would reproach someone who used them. But there will also be the intermediate kind of sentences, where we might allow their use."

Hence, the epistemic stance requires that agents make decisions on what is or is not appropriate to assert, based on their past experience of language use and on the assumption that there are existing linguistic conventions that should be adhered to if they do not wish to be misunderstood or contradicted. This decision problem would naturally lead agents to consider their subjective beliefs concerning the appropriateness of the available description labels in a given context. Uncertainty about such beliefs could then be quantified by using subjective probabilities, as proposed originally by de Finetti [3] and Ramsey [31] for other types of epistemic uncertainty. In the next section we introduce *label semantics* as a subjective probability based model of an agent's uncertainty about the appropriateness of expressions to describe a given instance, consistent with them adopting the epistemic stance.

To summarize, both the epistemic theory and the theory outlined in this paper identify vagueness as being a type of ignorance. For the epistemic theory this ignorance concerns the objective (but partially unknown) boundaries of concepts. While in our approach the focus is on an individual agent's ignorance of the underlying linguistic conventions governing the use of concept labels as part of communications between agents. In both cases, the association of vagueness with ignorance strongly contrasts with the many-valued/fuzzy logic approach in which the applicability of concept labels is viewed as being a matter of degree. Over recent years an extensive literature has emerged focusing on logical aspects of the fuzzy approach [14,28], including for example, embedding probability theory within a many-valued logic framework [9]. This work, however, is not within the scope of our paper.

3. The label semantics framework

Label semantics proposes two fundamental and inter-related measures of the appropriateness of labels as descriptions of an object or value. Given a finite set of labels LA a set of compound expressions LE can then be generated through recursive applications of logical connectives. The labels $L_i \in LA$ are intended to represent words such as adjectives and nouns which can be used to describe elements from the underlying universe Ω. In other words, L_i correspond to description labels for which the expression 'x is L_i' is meaningful for any $x \in \Omega$. This is exactly the type of expressions that Zadeh considers in his work in linguistic variables [40–42]. For example, if Ω is the set of all possible *rgb* values then LA could consist of the basic colour labels such as *red*, *yellow*, *green*, *orange* etc. In this case LE then contains those compound expression such as *red & yellow*, *not blue nor orange* etc. Note that in contrast to Kit Fine [8] in his discussion of penumbral connections we do not make the a priori assumption that the labels LA are mutually exclusive and exhaustive. This potentially allows for compound expressions such as *red & orange* to explicitly refer to boundaries between labels. The measure of appropriateness of an expression $\theta \in LE$ as a description of instance x is denoted by $\mu_\theta(x)$ and quantifies the agent's subjective belief that θ can be used to describe x based on his/her (partial) knowledge of the current labeling conventions of the population. From an alternative perspective, when faced with an object to describe, an agent may consider each label in LA and attempt to identify the subset of labels that are appropriate to use. Let this set be denoted by $\mathcal{D}_x$. In the face of their uncertainty regarding labeling conventions the agent will also be uncertain as to the composition of $\mathcal{D}_x$, and in label semantics this is quantified by a probability mass function $m_x : 2^{LA} \to [0,1]$ on subsets of labels. The relationship between these two measures will be described below.

Definition 1 *(Label expressions).* Given a finite set of labels LA the corresponding set of label expressions LE is defined recursively as follows:

- If $L \in LA$ then $L \in LE$.
- If $\theta, \varphi \in LE$ then $\neg\theta, \theta \wedge \varphi, \theta \vee \varphi \in LE$.

The mass function m_x on sets of labels then quantifies the agent's belief that any particular subset of labels contains all and only the labels with which it is appropriate to describe x i.e. $m_x(F)$ is the agent's subjective probability that $\mathcal{D}_x = F$.

Definition 2 *(Mass function on labels).* $\forall x \in \Omega$ a mass function on labels is a function $m_x : 2^{LA} \to [0,1]$ such that $\sum_{F \subseteq LA} m_x(F) = 1$.

The appropriateness measure, $\mu_\theta(x)$, and the mass function m_x are then related to each other on the basis that asserting 'x is θ' provides direct constraints on $\mathcal{D}_x$. For example, asserting 'x is L' for $L \in LA$ implies that L is appropriate to describe x and hence that $L \in \mathcal{D}_x$. Furthermore, asserting 'x is $L_1 \wedge L_2$', for labels $L_1, L_2 \in LA$ is taken as conveying the information that both L_1 and L_2 are appropriate to describe x so that $\{L_1, L_2\} \subseteq \mathcal{D}_x$. Similarly, '$x$ is $\neg L$' implies that L is not appropriate to describe x so $L \notin \mathcal{D}_x$. In general we can recursively define a mapping $\lambda : LE \to 2^{2^{LA}}$ from expressions to sets of subsets of labels, such that the assertion 'x is θ' directly implies the constraint $\mathcal{D}_x \in \lambda(\theta)$ and where $\lambda(\theta)$ is dependent on the logical structure of θ. For example, if $LA = \{low, medium, high\}$ then $\lambda(medium \wedge \neg high) = \{\{low, medium\}, \{medium\}\}$ corresponding to those sets of labels which include *medium* but do not include *high*.

Definition 3 *(λ-mapping).* $\lambda : LE \to 2^{2^{LA}}$ is defined recursively as follows: $\forall \theta, \varphi \in LE$

- $\forall L_i \in LA\ \lambda(L_i) = \{F \subseteq LA\colon L_i \in F\}$.
- $\lambda(\theta \wedge \varphi) = \lambda(\theta) \cap \lambda(\varphi)$.
- $\lambda(\theta \vee \varphi) = \lambda(\theta) \cup \lambda(\varphi)$.
- $\lambda(\neg\theta) = \lambda(\theta)^c$.

Based on the λ mapping we then define $\mu_\theta(x)$ as the sum of m_x over those set of labels in $\lambda(\theta)$.

Definition 4 *(Appropriateness measure).* The appropriateness measure defined by mass function m_x is a function $\mu : LA \times \Omega \to [0,1]$ satisfying

$$\forall \theta \in LE,\ \forall x \in \Omega \quad \mu_\theta(x) = \sum_{F \in \lambda(\theta)} m_x(F)$$

where $\mu_\theta(x)$ is used as shorthand notation for $\mu(\theta, x)$.

Note that in label semantics there is no requirement for the mass associated with the empty set to be zero. Instead, $m_x(\emptyset)$ quantifies the agent's belief that none of the labels are appropriate to describe x. We might observe that this phenomena occurs frequently in natural language, especially when labeling perceptions generated along some continuum. For example, we occasionally encounter colours for which none of our available colour descriptors seem appropriate. Hence, the value $m_x(\emptyset)$ is an indicator of the describability of x in terms of the labels LA.

Semantic relations $\models$ meaning 'more specific than' and $\equiv$ meaning 'equivalent to' can be defined on LE in the classical manner. Let Val be the set of valuation functions $v : LA \to \{0,1\}$ where for $L_i \in LA$, $v(L_i) = 1$ means that L_i is appropriate in the current context. In particular, the epistemic stance dictates that for each $x \in \Omega$ there would be a corresponding valuation v_x (partially unknown to the agent) determining which labels are appropriate to describe x. A valuation $v \in Val$ naturally determines an extension $v : LE \to \{0,1\}$ defined recursively as follows: For $\theta, \varphi \in LE$; $v(\theta \vee \varphi) = \max(v(\theta), v(\varphi))$, $v(\theta \wedge \varphi) = \min(v(\theta), v(\varphi))$, and $v(\neg\theta) = 1 - v(\theta)$. We can now define $\models$ and $\equiv$ as follows:

Definition 5. $\forall \theta, \varphi \in LE$

- $\theta \models \varphi$ if $\forall v \in Val\ v(\theta) = 1 \Rightarrow v(\varphi) = 1$.
- $\theta \equiv \varphi$ if $\forall v \in Val\ v(\theta) = v(\varphi)$.
- θ is a tautology if $\forall v \in Val\ v(\theta) = 1$.
- θ is a contradiction if $\forall v \in Val\ v(\theta) = 0$.

Given Definitions 4 and 5 it can be shown that appropriateness measures have the following general properties [21,22]:

Theorem 6 *(General properties of appropriateness measures). $\forall \theta, \varphi \in LE$ the following properties hold*:

- *If $\theta \models \varphi$ then $\forall x \in \Omega\ \mu_\theta(x) \leqslant \mu_\varphi(x)$.*
- *If $\theta \equiv \varphi$ then $\forall x \in \Omega\ \mu_\theta(x) = \mu_\varphi(x)$.*
- *If θ is a tautology then $\forall x \in \Omega\ \mu_\theta(x) = 1$.*
- *If θ is a contradiction then $\forall x \in \Omega\ \mu_\theta(x) = 0$.*
- *If $\theta \wedge \varphi$ is a contradiction then $\forall x \in \Omega\ \mu_{\theta\vee\varphi}(x) = \mu_\theta(x) + \mu_\varphi(x)$.*
- *$\forall x \in \Omega\ \mu_{\neg\theta}(x) = 1 - \mu_\theta(x)$.*
- *For $F \subseteq LA$ let $\theta_F = (\bigwedge_{L_i \in F} L_i) \wedge (\bigwedge_{L_i \notin F} \neg L_i)$ then $m_x(F) = \mu_{\theta_F}(x)$.*

From Theorem 6 it can be seen that for a fixed $x \in \Omega$, appropriateness measures correspond to probabilities on LE. More specifically, $\mu_\theta(x)$ can be interpreted as the subjective conditional probability that θ is an appropriate expression *given* that the element x is being described [21,22]. This naturally links label semantics to the conditional probability interpretation of fuzzy sets proposed by Hisdal [19] and widely developed by Coletti and Scozzafava [2]. However, as we will see in the following section, the mass function based definition of appropriateness measures enables us to explore links with random set theory and to formulate certain consonance assumptions in a straightforward manner. The random set approach is also well suited to exploring the link between label semantics and prototype theory as shown in Section 5.

3.1. Functionality of appropriateness measures

From Definition 4 we see that in order to be able to evaluate the appropriateness measure of any expression $\theta \in LE$ as a description for $x \in \Omega$ we must potentially know the value of m_x for all subsets of LA. Hence, we are, in principle, required to specify of order $2^{|LA|} - 1$ values for m_x. For large basic label sets this is clearly computationally infeasible. One solution to this problem would be to make additional assumptions about the definition of the mass assignment m_x so that there exists a functional relationship between the appropriateness measure for the basic labels (i.e. $\mu_L(x)$: $L \in LA$) and m_x. This would result in a functional calculus for appropriateness measures according to which the appropriateness of any compound expression could be determined directly from the appropriateness of the basic labels in the following sense:

Definition 7 *(Functional measures).* A measure $\mu : LE \times \Omega \to [0,1]$ is said to be functional if $\forall \theta \in LE$ there exists a function $f_\theta : [0,1]^{|LA|} \to [0,1]$ such that $\forall x \in \Omega\ \mu_\theta(x) = f_\theta(\mu_L(x)\colon\ L \in LA)$.

Now fuzzy logic is clearly functional in the sense of Definition 7 but it also satisfies the stronger property of truth-functionality. Truth-functionality defines the mapping f_θ to be a recursive combination of functions representing each connective, as determined by the logical structure of the expression θ. More formally:

Definition 8 *(Truth-functional measures).* A measure $\mu : LE \times \Omega \to [0,1]$ is said to truth-functional if there exists mappings $f_\neg : [0,1] \to [0,1]$, $f_\wedge : [0,1]^2 \to [0,1]$ and $f_\vee : [0,1]^2 \to [0,1]$ and such that $\forall \theta, \varphi \in LE$:

- $\forall x \in \Omega\ \mu_{\neg\theta}(x) = f_\neg(\mu_\theta(x))$.

- $\forall x \in \Omega \ \ \mu_{\theta \wedge \varphi}(x) = f_\wedge(\mu_\theta(x), \mu_\varphi(x))$.
- $\forall x \in \Omega \ \ \mu_{\theta \vee \varphi}(x) = f_\vee(\mu_\theta(x), \mu_\varphi(x))$.

Now from Theorem 6 it follows that appropriateness measures must satisfy the laws of excluded middle and idempotence. Hence, by the following theorem due to Dubois and Prade [4] they cannot be truth-functional except in the trivial case where all appropriateness values are either 0 or 1.

Theorem 9. *(See Dubois and Prade [4].) If $\mu : LE \times \Omega \to [0,1]$ is a truth-functional measure and satisfies both idempotence and the law of excluded middle then $\forall \theta \in LE$, $\forall x \in \Omega$, $\mu_\theta(x) \in \{0,1\}$.*

However, Theorem 9 does not apply to all functional measures, only those which are truth-functional, hence it may still be possible to define a functional calculus for appropriateness measures consistent with both Definitions 4 and 7. To investigate this possibility further we consider the relationship between appropriateness measures of compound expressions and those of the basic labels, imposed by Definition 4.

From Definition 3 the lambda mapping for a basic label $L_i \in LA$ is given by $\lambda(L_i) = \{F \subseteq LA\colon\ L_i \in F\}$ and hence the mass function m_x must satisfy the following constraint imposed by the appropriateness measures for the basic labels $\mu_{L_i}(x)\colon$ $L_i \in LA$:

$$\forall x \in \Omega,\ \forall L_i \in LA \quad \mu_{L_i}(x) = \sum_{F \subseteq LA\colon\ L_i \in F} m_x(F)$$

This constraint, however, is not sufficient to identify a unique mass function m_x given values for $\mu_{L_i}(x)\colon\ L_i \in LA$. Indeed, there are in general an infinite set of mass functions satisfying the above equation for a given set of basic label appropriateness values. Hence, in this context, the assumption of a functional calculus for appropriateness measures is equivalent to the assumption of a selection function which identifies a unique mass function from this set [21,22].

Definition 10 *(Selection function).* Let $\mathcal{M}$ be the set of all mass functions on 2^{LA}. Then a selection function is a function $\eta : [0,1]^{|LA|} \to \mathcal{M}$ such that if $\forall x \in \Omega\, \eta(\mu_{L_i}(x) : L_i \in LA) = m_x$ then

$$\forall x \in \Omega \ \ \forall L_i \in LA \qquad \sum_{F \subseteq LA\colon\ L_i \in F} m_x(F) = \mu_{L_i}(x)$$

Now since the value of $\mu_\theta(x)$ for any expression $\theta \in LE$ can be evaluated directly from m_x, then given a selection function η we have a functional method for determining $\mu_\theta(x)$ from the basic label appropriateness values, where f_θ in Definition 7 is given by:

$$f_\theta\big(\mu_{L_i}(x)\colon\ L_i \in LA\big) = \sum_{F \in \lambda(\theta)} \eta\big(\mu_{L_i}(x)\colon\ L_i \in LA\big)(F)$$

Two examples of selection functions are the consonant and the independent selection functions as defined below:

Definition 11 *(Consonant selection function).* Given non-zero appropriateness measures on basic labels $\mu_{L_i}(x)\colon\ i = 1, \ldots, n$ ordered such that $\mu_{L_i}(x) \geqslant \mu_{L_{i+1}}(x)$ for $i = 1, \ldots, n$ then the consonant selection function identifies the mass function,

$$m_x\big(\{L_1, \ldots, L_n\}\big) = \mu_{L_n}(x)$$
$$m_x\big(\{L_1, \ldots, L_i\}\big) = \mu_{L_i}(x) - \mu_{L_{i+1}}(x) \quad \text{for } i = 1, \ldots, n$$
$$m_x(\emptyset) = 1 - \mu_{L_1}(x)$$

Definition 12 *(Independent selection function).* Given appropriateness measures on basic labels $\mu_{L_i}(x)\colon\ L_i \in LA$ then the independent selection function identifies the mass function,

$$\forall F \subseteq LA \quad m_x(F) = \prod_{L_i \in F} \mu_{L_i}(x) \times \prod_{L_i \notin F} \big(1 - \mu_{L_i}(x)\big)$$

The consonant selection function corresponds to the assumption that for each $x \in \Omega$ an agent first identifies a total ordering on the appropriateness of labels. They then evaluate their belief values m_x about which labels are appropriate to describe x in such a way so as to be consistent with this ordering. More formally, let $\preceq_x$ denote the appropriateness ordering on LA for element x so that $L_1 \preceq_x L_2$ means that L_2 is at least as appropriate as L_1 for describing x. When evaluating $m_x(F)$ for $F \subseteq LA$ the agent then makes the assumption that the mass value is non-zero only if for every label $L_i \in F$ it also holds that $L_j \in F$ for every $L_j \in LA$ for which $L_i \preceq_x L_j$.

The independent selection function simply assumes that when judging the appropriateness of a label an agent does not take into account the level of appropriateness of any other label. Although this may seem difficult to justify, it could be reasonable in cases where labels relate to different facets of the object. For example, the appropriateness of the label *thin* might well be assumed to be independent of the appropriateness of the label *rich*.

The following theorems show that the consonant and independent selection functions result in familiar combination operators for restricted sets of expressions (see [21,22]).

Theorem 13. *(See [21,36].) Let $LE^{\wedge,\vee} \subseteq LE$ denote those expressions generated recursively from LA using only the connectives $\wedge$ and $\vee$. If $\forall x \in \Omega$, m_x is determined from $\mu_L(x)$: $L \in LA$ according to the consonant selection function then $\forall \theta, \varphi \in LE^{\wedge,\vee}$, $\forall x \in \Omega$ it holds that*:

$$\mu_{\theta\wedge\varphi}(x) = \min\bigl(\mu_\theta(x), \mu_\varphi(x)\bigr) \quad \textit{and} \quad \mu_{\theta\vee\varphi}(x) = \max\bigl(\mu_\theta(x), \mu_\varphi(x)\bigr)$$

Theorem 14. *(See [22].) If $\forall x \in \Omega$, m_x is determined from $\mu_L(x)$: $L \in LA$ according to the independent selection function then for labels $L_1, \ldots, L_n \in LA$ we have that $\forall x \in \Omega$:*

$$\mu_{L_1\wedge L_2\wedge\cdots\wedge L_n}(x) = \prod_{i=1}^{n} \mu_{L_i}(x)$$

$$\mu_{L_1\vee L_2\vee\cdots\vee L_n}(x) = \sum_{\emptyset\neq S\subseteq\{L_1,\ldots,L_n\}} (-1)^{|S|-1} \prod_{L_i\in S} \mu_{L_i}(x)$$

One interpretation of selection functions is that they provide a means of encoding the semantic dependence between labels. From this perspective, the consonant selection function assumes that the appropriateness of all labels are assessed on the basis of the same set of shared attributes i.e. that they can be represented within a single conceptual space [11]. Typical examples might be height labels such as tall, medium and short or colour labels. Alternatively, the independent selection function assumes a set of labels where the appropriateness of each label is judged on the basis of a set of attributes independent from those employed to assess the appropriateness of any other label i.e. we have a set of independent conceptual spaces (one for each label) with no shared or dependent attributes. Intermediate cases between the consonant and independent selection functions can also be considered, perhaps resulting in the kind of partial orderings on appropriateness proposed in [23]. We shall return to the discussion of semantic dependence in the sequel where we will consider the issue from a prototype theory perspective.

3.2. Assertability decisions

For any given $x \in \Omega$ and expression $\theta \in LE$ it remains unclear exactly how an agent would use their evaluation of the appropriateness measure $\mu_\theta(x)$ in order to reach a decision as to whether or not the statement 'x is θ' is assertible. One possibility would be to use a threshold based approach according to which a positive decision to assert 'x is θ' would require that the appropriateness measure $\mu_\theta(x)$ be sufficiently close to 1. More formally, an agent would be willing to assert 'x is θ' at certainty level $\alpha \geqslant 0.5$, denoted $Assert_\alpha(x \text{ is } \theta)$, if $\mu_\theta(x) \geqslant \alpha$. For a tautology $\theta \vee \neg\theta$ we would then have $Assert_\alpha(x \text{ is } \theta \vee \neg\theta)$ holding at any certainty level α. However, for $\alpha \geqslant \max(\mu_\theta(x), 1 - \mu_\theta(x))$ neither 'x is θ' nor 'x is $\neg\theta$' is assertible at certainty level α i.e. $Assert_\alpha(x \text{ is } \theta) \vee Assert_\alpha(x \text{ is } \neg\theta)$ does not hold. This is consistent with the intuition that, for example, an agent would happily concede that any given colour is either red or not red even though for certain borderline cases they would be unwilling to use either the description *red* or the description *not red*. In general, for $x \in \Omega$ and expressions $\theta, \varphi \in LE$ $Assert_\alpha(x \text{ is } \theta \vee \varphi)$ is not equivalent to $Assert_\alpha(x \text{ is } \theta) \vee Assert_\alpha(x \text{ is } \varphi)$ since the former requires that $\mu_{\theta\vee\varphi}(x) \geqslant \alpha$ while the latter requires that $\max(\mu_\theta(x), \mu_\varphi(x)) \geqslant \alpha$. Indeed, if the consonant selection function is applied it can be seen from Theorem 13 that these two assertability statements are only equivalent for expressions θ and φ which do not involve negation i.e. $\theta, \varphi \in LE^{\wedge,\vee}$.

Notice that although $\mu_\theta(x)$ is an indication of the assertability of the statement 'x is θ' it should not be interpreted as the probability that this assertion will actually occur during communications aimed at describing x. Instead, $\mu_\theta(x)$ quantifies the belief that the expression θ is appropriate to describe x, where appropriateness is judged purely on the basis of the agent's interpretation of the meaning of θ. This understanding of meaning would in turn be based on an internal representation[2] of the labels in *LA*, as inferred from the agent's past experience of communications involving these labels. This is not the same as taking $\mu_\theta(x)$ as corresponding to the probability of assertion 'x is θ' actually occurring. For example, given a basic label $L_i \in LA$ then from a semantic perspective the expressions L_i and $\neg\neg L_i$ are equally appropriate to describe any given x (with equal appropriateness measures). However, for those x with high appropriateness measures for these equivalent expressions, the assertion 'x is L_i' is perhaps much more likely to be actually used than the assertion 'x is $\neg\neg L_i$', for reasons of syntactic simplicity. Indeed, Lawry [24] proposes a model for evaluating the probability of assertions, based on

[2] In the sequel we shall suggest that this internal representation might be based on prototypes.

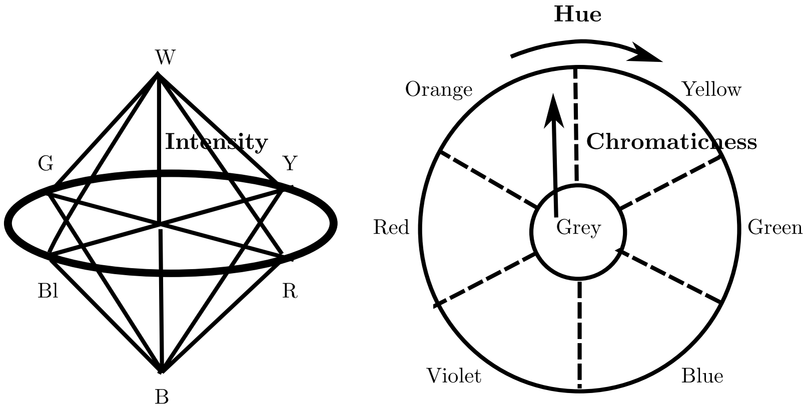

Fig. 1. NCS colour spindle [35]: A conceptual space for representing colour categories with dependent attributes Hue, Chromaticness and Intensity.

appropriateness measures, which takes into account a prior weighting on expressions dependent, for example, on their relative syntactic complexity. This, however, is beyond the scope of the current paper.

4. Prototype theory and vagueness

The central tenet of Prototype theory (Rosch [32,33]) is that concepts, rather than being defined by formal rules or mappings, are represented by prototypes and that categorization is based on similarity to these prototypes (see Hampton [15] for an overview). By taking typicality to be a decreasing function of distance from prototypes, this approach would naturally explain the fact that some instances are seen as being more typical exemplars of a concept than others. For example, robins are more typical exemplars of birds than penguins, since the latter have certain atypical characteristics such as the inability to fly. This notion of typicality is also strongly related to concept vagueness where borderline cases have an intermediate range of typicality values. In other words, such cases are not sufficiently similar to the concept prototypes to be judged as having certain membership in the category but are also not sufficiently dissimilar to the prototypes to be ruled as being certainly outside the category.

Gärdenfors [11] has recently introduced the notion of conceptual spaces for concept representation, corresponding to a metric space of (possibly) dependent attributes or features of elements from the underlying universe Ω. For example, the NCS colour spindle [35] is a proposed conceptual space for colour categories based on polar attributes *Hue* and *Chromaticness* and where Chromaticness is constrained by a third attribute, *Intensity* (see Fig. 1). Within a conceptual space *properties* are represented by convex regions. This provides a natural link to prototype theory since given a convex set of prototypes for each property the space is partitioned into convex regions each defined as the set of points closest to the prototypes for a given property. Such partitions are referred to as Voronoi tessellations.

Similarity to prototypes has been widely suggested as a possible basis for membership functions in fuzzy logic. Dubois and Prade [6] and also Dubois et al. [7] identify such an interpretation as one of the three main semantics for membership degrees. For example, Ruspini [34] introduces a semantics for fuzzy reasoning where, given a measure of similarity between elements with a range between 0 (totally dissimilar) and 1 (totally similar), the membership degree of an element x in a concept corresponds to the supremum of the similarity values between x and the prototypes for the concept. Hampton [16] proposed a thresholding model for categorization whereby an instance is positively classified as belonging to a category if its similarity to the prototypes of that category exceeds a certain threshold. In the sequel we demonstrate the relationship between label semantics and exactly such a thresholding model.

5. A prototype theory interpretation of label semantics

In this interpretation it is proposed that the basic labels in *LA* correspond to natural categories each with an associated set of prototypes. A label L is then deemed to be an appropriate description of an element $x \in \Omega$ provided that x is *sufficiently similar* to the prototypes for L. The requirement of being 'sufficiently similar' is clearly imprecise and is modelled here by introducing an uncertain threshold on distance from prototypes. In keeping with the epistemic stance this uncertainty is assumed to be probabilistic in nature. In other words, an agent believes that there is some optimal threshold of this kind according to which he or she is best able to abide by the conventions of language when judging the appropriateness of labels. However, the agent is uncertain as to exactly what this threshold should be and instead defines a probability distribution on potential threshold values. Notice that the idea of an uncertain threshold fits well with the epistemic theory of vagueness since it naturally results in uncertain (but crisp) concept boundaries. Although, as discussed in Section 2, the assumption of an unknown objective definition of the concept is not required. Instead within such a model the agent's knowledge of the threshold probability distribution and also of the label prototypes would be derived from their experience of language use across a population of communicating agents each adopting the epistemic stance.

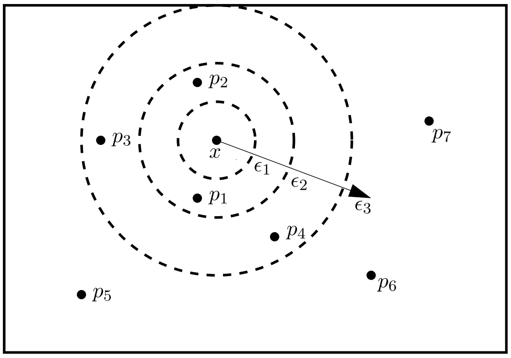

Fig. 2. Identifying $\mathcal{D}_x^{\epsilon}$ as ϵ varies, where $P_i = \{p_i\}$ for $L_i \in LA$ are singletons; For ϵ_1, ϵ_2 and ϵ_3 shown in the diagram $\mathcal{D}_x^{\epsilon_1} = \emptyset$, $\mathcal{D}_x^{\epsilon_2} = \{L_1, L_2\}$, $\mathcal{D}_x^{\epsilon_3} = \{L_1, L_2, L_3, L_4\}$.

In this section we begin by introducing a special case of the prototype model that results in a calculus for appropriateness measures consistent with the consonant selection function (Definition 11). We then propose a more general theory which can represent a range of semantic dependencies between labels.

5.1. A consonant model

Suppose that a distance function[3] d is defined on Ω such that $d : \Omega^2 \to [0, \infty)$ and satisfies $d(x, x) = 0$ and $d(x, y) = d(y, x)$ for all elements $x, y \in \Omega$. This function is then extended to sets of elements such that for $S, T \subseteq \Omega$, $d(S, T) = \inf\{d(x, y)\colon\ x \in S \text{ and } y \in T\}$. For each label $L_i \in LA$ let there be a set $P_i \subseteq \Omega$ corresponding to prototypical elements for which L_i is certainly an appropriate description. Within this framework L_i is deemed to be appropriate to describe an element $x \in \Omega$ provided x is sufficiently close or similar to a prototypical element in P_i. This is formalized by the requirement that x is within a maximal distance threshold ϵ of P_i. I.e. L_i is appropriate to describe x if $d(x, P_i) \leqslant \epsilon$ where $\epsilon \geqslant 0$. From this perspective an agent's uncertainty regarding the appropriateness of a label to describe a value x is characterised by his or her uncertainty regarding the distance threshold ϵ. In fact, it is also possible that an agent could be uncertain regarding the definition of the prototype sets P_i. However, for this paper we make the simplification assumption that no uncertainty is associated with the prototypes. Here we assume that ϵ is a random variable and that the uncertainty is represented by a probability density function δ for ϵ defined on $[0, \infty)$.[4] Within this interpretation a natural definition of the complete description of an element $\mathcal{D}_x$ and the associated mass function m_x can be given as follows:

Definition 15. For $\epsilon \in [0, \infty)$ $\mathcal{D}_x^{\epsilon} = \{L_i \in LA\colon\ d(x, P_i) \leqslant \epsilon\}$ and $\forall F \subseteq LA$ $m_x(F) = \delta(\{\epsilon\colon\ \mathcal{D}_x^{\epsilon} = F\})$.[5]

Intuitively speaking $\mathcal{D}_x^{\epsilon}$ identifies the set of labels with prototypes lying within ϵ of x. Fig. 2 shows $\mathcal{D}_x^{\epsilon}$ in a hypothetical conceptual space as ϵ varies. Notice that the sequence $\mathcal{D}_x^{\epsilon}$ as ϵ varies generates a nested hierarchy of label sets. Furthermore, the distance metric d naturally generates a total ordering on the appropriateness of labels for any element x, according to which label L_j is as least as appropriate to describe x as label L_i if x is closer (or equidistant) to P_j than to P_i i.e. $L_i \preceq_x L_j$ iff $d(x, P_i) \geqslant d(x, P_j)$ as suggested in Section 3.1.

Also notice from Definition 15, that for $L_i \in LA$ the appropriateness measure $\mu_{L_i}(x)$ is given by $\delta(\{\epsilon\colon\ L_i \in \mathcal{D}_x^{\epsilon}\})$. Consequently, if we view $\mathcal{D}_x^{\epsilon}$ as a random set from $[0, \infty)$ into 2^{LA} then $\mu_{L_i}(x)$ corresponds to the single point coverage function[6] of $\mathcal{D}_x^{\epsilon}$. This provides us with a link to the random set interpretation of fuzzy sets (see [6,7,12,13,27]) except that in this case the random set maps to sets of labels rather than sets of elements. Hence, the interpretation of label semantics as proposed above provides a link between random set theory and prototype theory. A further consequence of this relationship with random set theory is that $\forall x \in \Omega$, the mass function m_x must satisfy the conditions given in Definition 11. This follows from the well-known fact that the mass function of a nested (or consonant) finite random set, in this case $\mathcal{D}_x^{\epsilon}$, can be determined uniquely from its single point coverage function (see [22] for an exposition).

The following results show how the appropriateness of an expression $\theta \in LE$ to describe an element x is equivalent to a constraint $\epsilon \in I(\theta, x)$, for a measurable subset $I(\theta, x)$ of $[0, \infty)$ defined as follows:

---

[3] d may also be a distance metric if it satisfies the triangular inequality $\forall x, y, z \in \Omega$ $d(x, z) \leqslant d(x, y) + d(y, z)$, but this is not strictly required.

[4] Even though ϵ is a random variable there is no suggestion of any underlying stochastic process. Instead an agent's uncertainty concerning ϵ is epistemic in nature and dependent on their experience of language use as discussed in Section 2.1.

[5] For Lebesgue measurable set I, we denote $\delta(I) = \int_I \delta(\epsilon)\, d\epsilon$ i.e. we also use δ to denote the probability measure induced by density function δ.

[6] For a finite random set $\mathcal{S}$ into $2^{\mathcal{U}}$, where $\mathcal{U}$ is the underlying finite universe, the associate mass function $m : 2^{\mathcal{U}} \to [0, 1]$ is such that for $T \subseteq \mathcal{U}$, $m(T)$ is the probability that $\mathcal{S} = T$. The single point coverage function then corresponds to the probability that $z \in \mathcal{S}$, for $z \in \mathcal{U}$, and is given by $\sum_{T\colon\ z \in T} m(T)$. In label semantics $\mathcal{U} = LA$ and $\mathcal{S} = \mathcal{D}_x^{\epsilon}$.

Definition 16. $\forall x \in \Omega$ and $\theta \in LE$, $I(\theta, x) \subseteq [0, \infty)$ is defined recursively as follows: $\forall \theta, \varphi \in LE$

- $\forall L_i \in LA$ $I(L_i, x) = [d(x, P_i), \infty)$.
- $I(\theta \wedge \varphi, x) = I(\theta, x) \cap I(\varphi, x)$.
- $I(\theta \vee \varphi, x) = I(\theta, x) \cup I(\varphi, x)$.
- $I(\neg\theta, x) = I(\theta, x)^c$.

Theorem 17.

$$\forall \theta \in LE,\ \forall x \in \Omega \quad I(\theta, x) = \left\{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\theta)\right\}$$

Proof. Let

$$LE^{(1)} = LA \quad \text{and} \quad LE^{(k)} = LE^{(k-1)} \cup \left\{\theta \wedge \varphi, \theta \vee \varphi, \neg\theta\colon \theta, \varphi \in LE^{(k-1)}\right\}$$

We now prove the result by induction on k.

Limit Case: $k = 1$ For $L_i \in LA$ we have that

$$I(L_i, x) = \left[d(x, P_i), \infty\right) = \left\{\epsilon\colon d(x, P_i) \leqslant \epsilon\right\} = \left\{\epsilon\colon L_i \in \mathcal{D}_x^\epsilon\right\} = \left\{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(L_i)\right\}$$

Inductive Step: Assume true for k For $\Phi \in LE^{(k+1)}$ either $\Phi \in LE^{(k)}$, in which case the result holds trivially by the inductive hypothesis, or one of the following holds:

- $\Phi = \theta \wedge \varphi$ so that $I(\Phi, x) = I(\theta \wedge \varphi, x) = I(\theta, x) \cap I(\varphi, x) = \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\theta)\} \cap \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\varphi)\}$ (by the inductive hypothesis) $= \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\theta) \cap \lambda(\varphi)\} = \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\theta \wedge \varphi)\} = \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\Phi)\}$.
- $\Phi = \theta \vee \varphi$ so that $I(\Phi, x) = I(\theta \vee \varphi, x) = I(\theta, x) \cup I(\varphi, x) = \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\theta)\} \cup \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\varphi)\}$ (by the inductive hypothesis) $= \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\theta) \cup \lambda(\varphi)\} = \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\theta \vee \varphi)\} = \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\Phi)\}$.
- $\Phi = \neg\theta$ so that $I(\Phi, x) = I(\neg\theta, x) = I(\theta, x)^c = \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\theta)\}^c = \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\theta)^c\} = \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\neg\theta)\} = \{\epsilon\colon \mathcal{D}_x^\epsilon \in \lambda(\Phi)\}$. □

Corollary 18.

$$\forall \theta \in LE,\ \forall x \in \Omega \quad \mu_\theta(x) = \delta\left(I(\theta, x)\right)$$

Proof. The result follows trivially from Theorem 17 and Definition 15. □

Furthermore, in the case that we restrict ourselves to expressions in $LE^{\wedge,\vee}$ (Theorem 13) the following result shows that $I(\theta, x)$ simply identifies a lower bound on ϵ.

Definition 19. We define $lb : LE^{\wedge,\vee} \times \Omega \to [0, \infty)$ recursively as follows: $\forall x \in \Omega$, $\forall \theta, \varphi \in LE^{\wedge,\vee}$

- $\forall L_i \in LA$ $lb(L_i, x) = d(x, P_i)$.
- $lb(\theta \wedge \varphi, x) = \max(lb(\theta, x), lb(\varphi, x))$ and $lb(\theta \vee \varphi, x) = \min(lb(\theta, x), lb(\varphi, x))$.

Theorem 20. *$\forall x \in \Omega$, $\forall x \in LE^{\wedge,\vee}$, then $I(\theta, x) = [lb(\theta, x), \infty)$.*

Proof. Let

$$LE_1^{\wedge,\vee} = LA \quad \text{and} \quad LE_k^{\wedge,\vee} = LE_{k-1}^{\wedge,\vee} \cup \left\{\theta \wedge \varphi,\ \theta \vee \varphi\colon \theta, \varphi \in LE_{k-1}^{\wedge,\vee}\right\}$$

We now prove the result by induction on k.

Limit Case: $k = 1$ For $L_i \in LA$ $I(L_i, x) = [d(x, P_i), \infty) = [lb(L_i, x), \infty)$ as required.

Inductive Step: Assume true for k For $\Phi \in LE_{k+1}^{\wedge,\vee}$ either $\Phi \in LE_k^{\wedge,\vee}$ in which case the result follows trivially from the inductive hypothesis or one of the following holds:

- $\Phi = \theta \wedge \varphi$ where $\theta, \varphi \in LE_k^{\wedge,\vee}$. In this case $I(\Phi, x) = I(\theta, x) \cap I(\varphi, x) = [lb(\theta, x), \infty) \cap [lb(\varphi, x), \infty)$ (by the inductive hypothesis)$= [\max(lb(\theta, x), lb(\varphi, x)), \infty) = [lb(\theta \wedge \varphi, x), \infty)$ as required.
- $\Phi = \theta \vee \varphi$ where $\theta, \varphi \in LE_k^{\wedge,\vee}$. In this case $I(\Phi, x) = I(\theta, x) \cup I(\varphi, x) = [lb(\theta, x), \infty) \cup [lb(\varphi, x), \infty)$ (by the inductive hypothesis)$= [\min(lb(\theta, x), lb(\varphi, x)), \infty) = [lb(\theta \vee \varphi, x), \infty)$ as required. □

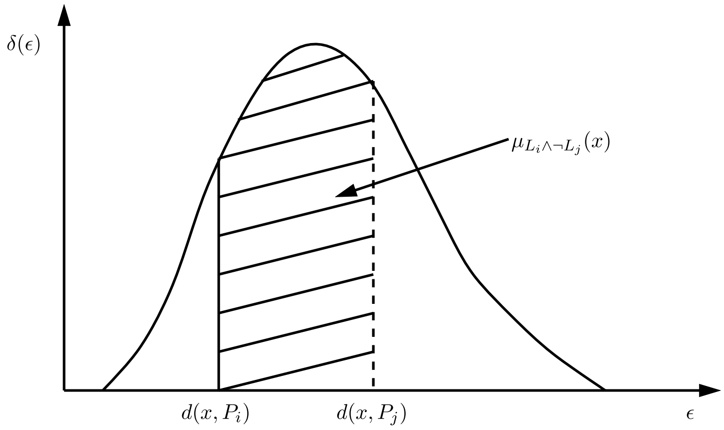

Fig. 3. Area under the density function δ corresponding to $\mu_{L_i \wedge \neg L_j}(x)$.

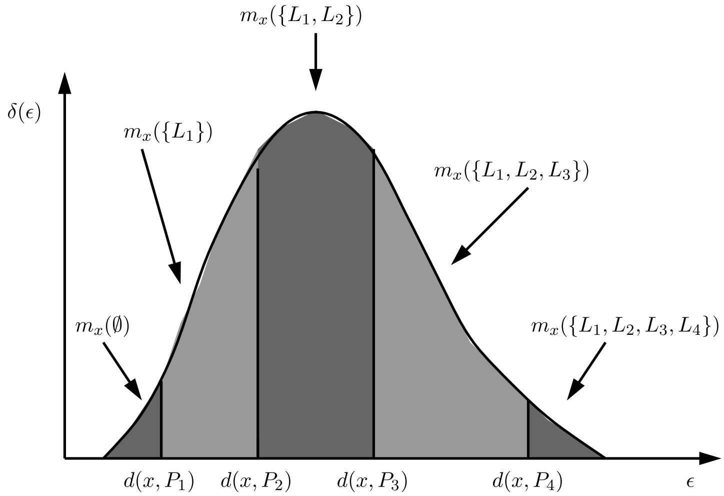

Fig. 4. Let $LA = \{L_1, L_2, L_3, L_4\}$ and $L_4 \preceq_x L_3 \preceq_x L_2 \preceq_x L_1$. This figure shows the values of m_x as areas under δ.

Notice, from Theorem 20 we have that $\forall \theta, \varphi \in LE^{\wedge,\vee}$ $\mu_{\theta\vee\varphi}(x) = \delta([lb(\theta \vee \varphi, x), \infty)) = \delta([\min(lb(\theta, x), lb(\varphi, x)), \infty)) = \max(\delta([lb(\theta, x), \infty)), \delta([lb(\varphi, x), \infty))) = \max(\mu_\theta(x), \mu_\varphi(x))$. Similarly, $\mu_{\theta\wedge\varphi}(x) = \min(\mu_\theta(x), \mu_\varphi(x))$ as is consistent with Theorem 13.

Example 21.

$$I(L_i, x) = \big[d(x, P_i), \infty\big), \qquad I(\neg L_i, x) = \big[0, d(x, P_i)\big), \qquad I(L_i \wedge L_j, x) = \big[\max\big(d(x, P_i), d(x, P_j)\big), \infty\big)$$
$$I(L_i \vee L_j, x) = \big[\min\big(d(x, P_i), d(x, P_j)\big), \infty\big)$$
$$I(L_i \wedge \neg L_j, x) = \begin{cases} [d(x, P_i), d(x, P_j)) & \text{if } d(x, P_i) < d(x, P_j) \\ \emptyset & \text{otherwise} \end{cases}$$

From Theorem 6 we have that for $F \subseteq LA$ $m_x(F) = \mu_{\theta_F}(x)$ where $\theta_F = (\bigwedge_{L\in F} L) \wedge (\bigwedge_{L\notin F} \neg L)$. Hence, $m_x(F) = \delta(I(\theta_F, x))$ where $I(\theta_F, x) = [\max\{d(x, P_i)\colon\ L_i \in F\}, \min\{d(x, P_i)\colon\ L_i \notin F\})$ provided that $\max\{d(x, P_i)\colon\ L_i \in F\} < \min\{d(x, P_i)\colon\ L_i \notin F\}$ and $= \emptyset$ otherwise.

Figs. 3 and 4 show the areas under δ corresponding to $\mu_{L_i\wedge\neg L_j}(x)$ and the values of the mass function m_x respectively.

Example 22. Let $\Omega = [0, 10]$ and the labels $LA = \{L_1, L_2, L_3\}$ be defined such that $P_1 = \{4\}$, $P_2 = [5, 7]$ and $P_3 = [7.5, 8]$ and let $d(x, y) = \|x - y\|$. Also let δ be a Gaussian density with mean 0 and standard deviation 0.6 renormalised so that area under [0, 10] is 1 (see Fig. 5). Now consider the element $x = 4.5$ then $\mathcal{D}^{\epsilon}_{4.5}$ is defined as follows:

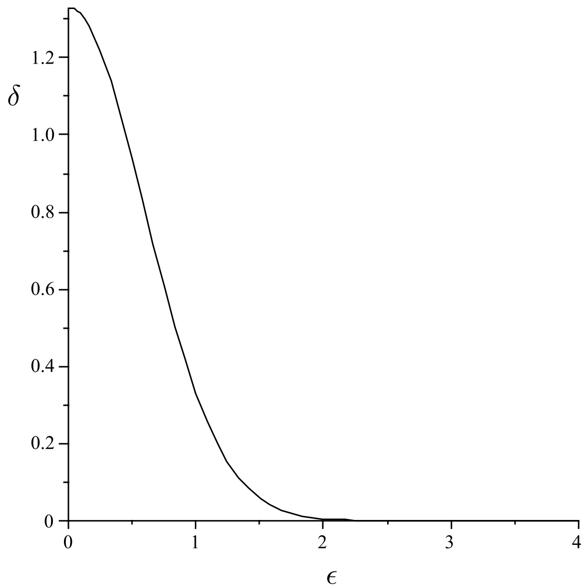

Fig. 5. Density function δ, normalised from a Gaussian with mean 0 and standard deviation 0.6.

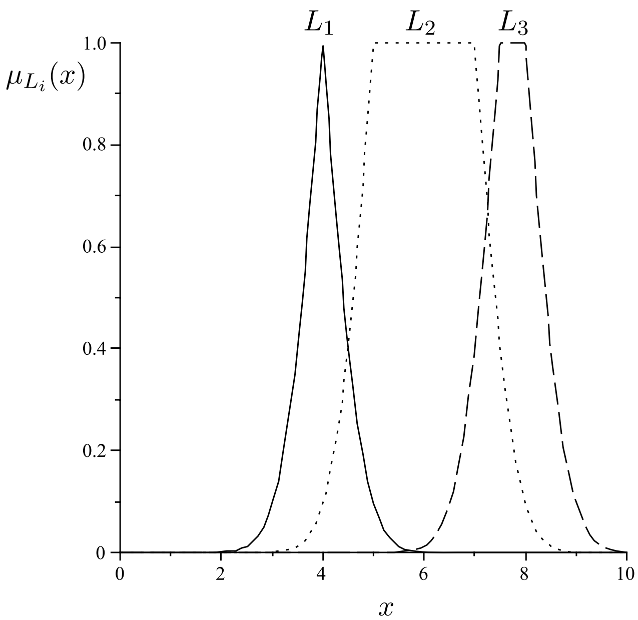

Fig. 6. Appropriateness measures for L_1, L_2 and L_3 derived from a Gaussian density δ and prototypes $P_1 = \{4\}$, $P_2 = [5, 7]$ and $P_3 = [7.5, 8]$.

$$\mathcal{D}^{\epsilon}_{4.5} = \begin{cases} \emptyset & \epsilon < 0.5 \\ \{L_1, L_2\} & 0.5 \leqslant \epsilon < 3 \\ \{L_1, L_2, L_3\} & \epsilon \geqslant 3 \end{cases}$$

The associated mass function $m_{4.5}$ is then given by

$$m_{4.5} = \emptyset\colon\ \delta\big([0, 0.5)\big) \approx 0.5953, \qquad \{L_1, L_2\}\colon\ \delta\big([0.5, 3)\big) \approx 0.4047, \qquad \{L_1, L_2, L_3\}\colon\ \delta\big([3, 10]\big) \approx 0$$

Fig. 6 shows the appropriateness measures for labels L_1, L_2 and L_3 corresponding to

$$\mu_{L_i}(x) = \delta\big(\big[d(x, P_i), \infty\big)\big) = \delta\big(\big[d(x, P_i), 10\big]\big)$$

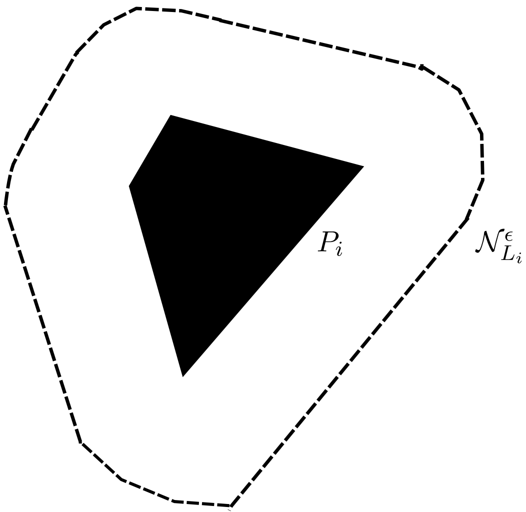

Fig. 7. $\mathcal{N}^{\epsilon}_{L_i}$ as a neighbourhood of P_i.

5.2. Neighbourhoods of prototypes

Another perspective on the prototype theory view of label semantics results from considering the set of elements from Ω which can be appropriately described by a label or expression. For an expression $\theta \in LE$ this subset of Ω would, in effect, correspond to the extension of the concept represented by θ. Now for the model proposed in Section 5.1 the set of elements which can be appropriately described by a label $L_i \in LA$ corresponds to the neighbourhood of P_i (see Fig. 7) defined by:

$$\mathcal{N}^{\epsilon}_{L_i} = \left\{x \in \Omega \colon d(x, P_i) \leqslant \epsilon\right\}$$

This can be naturally extended to any label expression $\theta \in LE$ as follows:

Definition 23.

$$\forall \theta \in LE \quad \mathcal{N}^{\epsilon}_{\theta} = \left\{x \in \Omega \colon \mathcal{D}^{\epsilon}_x \in \lambda(\theta)\right\} = \left\{x \in \Omega \colon \epsilon \in I(\theta, x)\right\}$$

The following theorem shows that $\mathcal{N}^{\epsilon}_{\theta}$ can be determined recursively from neighbourhoods $\mathcal{N}^{\epsilon}_{L_i}$ for the basic labels $L_i \in LA$.

Theorem 24. $\forall x \in \Omega,\ \forall \theta, \varphi \in LE,\ \forall \epsilon \in [0, \infty)$ *the following hold*:

(i) $\mathcal{N}^{\epsilon}_{\theta \wedge \varphi} = \mathcal{N}^{\epsilon}_{\theta} \cap \mathcal{N}^{\epsilon}_{\varphi}$.
(ii) $\mathcal{N}^{\epsilon}_{\theta \vee \varphi} = \mathcal{N}^{\epsilon}_{\theta} \cup \mathcal{N}^{\epsilon}_{\varphi}$.
(iii) $\mathcal{N}^{\epsilon}_{\neg\theta} = (\mathcal{N}^{\epsilon}_{\theta})^c$.

Proof.

(i) $\mathcal{N}^{\epsilon}_{\theta \wedge \varphi} = \{x\colon \mathcal{D}^{\epsilon}_x \in \lambda(\theta \wedge \varphi)\} =$ (by Definition 3) $\{x\colon \mathcal{D}^{\epsilon}_x \in \lambda(\theta) \cap \lambda(\varphi)\} = \{x\colon \mathcal{D}^{\epsilon}_x \in \lambda(\theta)\} \cap \{x\colon \mathcal{D}^{\epsilon}_x \in \lambda(\varphi)\} = \mathcal{N}^{\epsilon}_{\theta} \cap \mathcal{N}^{\epsilon}_{\varphi}$ (by Definition 23).
(ii) $\mathcal{N}^{\epsilon}_{\theta \vee \varphi} = \{x\colon \mathcal{D}^{\epsilon}_x \in \lambda(\theta \vee \varphi)\} =$ (by Definition 3) $\{x\colon \mathcal{D}^{\epsilon}_x \in \lambda(\theta) \cup \lambda(\varphi)\} = \{x\colon \mathcal{D}^{\epsilon}_x \in \lambda(\theta)\} \cup \{x\colon \mathcal{D}^{\epsilon}_x \in \lambda(\varphi)\} = \mathcal{N}^{\epsilon}_{\theta} \cup \mathcal{N}^{\epsilon}_{\varphi}$ (by Definition 23).
(iii) $\mathcal{N}^{\epsilon}_{\neg\theta} = \{x\colon \mathcal{D}^{\epsilon}_x \in \lambda(\neg\theta)\} =$ (by Definition 3) $\{x\colon \mathcal{D}^{\epsilon}_x \in \lambda(\theta)^c\} = \{x\colon \mathcal{D}^{\epsilon}_x \in \lambda(\theta)\}^c = (\mathcal{N}^{\epsilon}_{\theta})^c$ (by Definition 23). $\square$

Theorem 25.

$$\forall \theta \in LE,\ \forall x \in \Omega \quad \mu_{\theta}(x) = \delta\left(\left\{\epsilon\colon x \in \mathcal{N}^{\epsilon}_{\theta}\right\}\right)$$

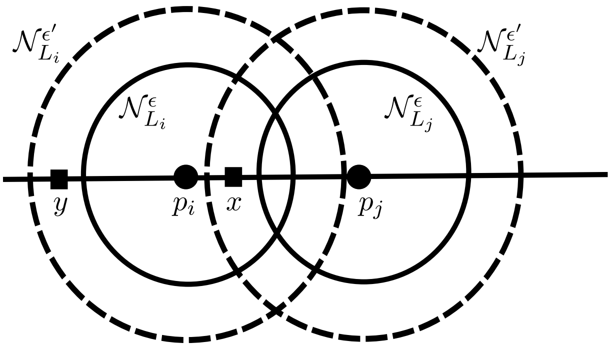

Fig. 8. For $P_i = \{p_i\}$ and $P_j = \{p_j\}$ x and y are elements such that $x \in N^{\epsilon}_{L_i \wedge \neg L_j}$ but $x \notin N^{\epsilon'}_{L_i \wedge \neg L_j}$, and $y \notin N^{\epsilon}_{L_i \wedge \neg L_j}$ but $y \in N^{\epsilon'}_{L_i \wedge \neg L_j}$. Hence $N^{\epsilon}_{L_i \wedge \neg L_j} \nsubseteq N^{\epsilon'}_{L_i \wedge \neg L_j}$ and $N^{\epsilon'}_{L_i \wedge \neg L_j} \nsubseteq N^{\epsilon}_{L_i \wedge \neg L_j}$.

Proof.

$$\begin{aligned}\delta\big(\big\{\epsilon\colon x \in \mathcal{N}^{\epsilon}_{\theta}\big\}\big) &= \text{(by Definition 23) } \delta\big(\big\{\epsilon\colon \mathcal{D}^{\epsilon}_{x} \in \lambda(\theta)\big\}\big) = \sum_{F\colon F \in \lambda(\theta)} \delta\big(\big\{\epsilon\colon \mathcal{D}^{\epsilon}_{x} = F\big\}\big) \\ &= \text{(by Definition 15) } \sum_{F\colon F \in \lambda(\theta)} m_x(F) = \mu_\theta(x) \qquad \square\end{aligned}$$

Theorem 25 provides us with an alternative characterisation of appropriateness measure $\mu_\theta(x)$ as the single point coverage function of the random set neighbourhood $\mathcal{N}^{\epsilon}_{\theta}$. Again this shows a link with the work of Goodman and Nguyen (see [12,13] and [27]) and their proposed interpretation of fuzzy sets as single point coverage functions of random sets.

The following theorem shows that for expressions θ not involving negation the random set neighbourhood $\mathcal{N}^{\epsilon}_{\theta}$ is nested.

Theorem 26. *$\forall \theta \in LE^{\wedge,\vee}$ and $\forall \epsilon' \geqslant \epsilon \geqslant 0$ $\mathcal{N}^{\epsilon}_{\theta} \subseteq \mathcal{N}^{\epsilon'}_{\theta}$.*

Proof.

$$\begin{aligned}\mathcal{N}^{\epsilon}_{\theta} &= \big\{x\colon \mathcal{D}^{\epsilon}_{x} \in \lambda(\theta)\big\} = \text{(by Theorem 17) } \big\{x\colon \epsilon \in I(\theta, x)\big\} \\ &= \text{(by Theorem 20) } \big\{x\colon lb(\theta, x) \leqslant \epsilon\big\} \subseteq \big\{x\colon lb(\theta, x) \leqslant \epsilon'\big\} \\ &= \big\{x\colon \epsilon' \in I(\theta, x)\big\} = \mathcal{N}^{\epsilon'}_{\theta} \text{ (by Definition 23)} \qquad \square\end{aligned}$$

For expressions $\theta \notin LE^{\wedge,\vee}$ it is not in general the case that $\mathcal{N}^{\epsilon}_{\theta}$ forms a nested sequence as ϵ varies. For example, consider the expression $L_i \wedge \neg L_j$ and suppose there are elements $x, y \in \Omega$ together with real values $\epsilon' > \epsilon > 0$ such that $d(x, P_i) \leqslant \epsilon < \min(d(x, P_j), d(y, P_i))$ and $\max(d(y, P_i), d(x, P_j)) \leqslant \epsilon' < d(y, P_j)$ (see for example Fig. 8). In this case $x \in \mathcal{N}^{\epsilon}_{L_i}$ and $x \notin \mathcal{N}^{\epsilon}_{L_j}$, so that $x \in \mathcal{N}^{\epsilon}_{L_i \wedge \neg L_j}$, while $y \notin \mathcal{N}^{\epsilon}_{L_i}$ so that $y \notin \mathcal{N}^{\epsilon}_{L_i \wedge \neg L_j}$. On the other hand $x \in \mathcal{N}^{\epsilon'}_{L_j}$ so that $x \notin \mathcal{N}^{\epsilon'}_{L_i \wedge L_j}$, while $y \in \mathcal{N}^{\epsilon'}_{L_i}$ and $y \notin \mathcal{N}^{\epsilon'}_{L_j}$ so that $y \in \mathcal{N}^{\epsilon'}_{L_i \wedge L_j}$. Hence, $\mathcal{N}^{\epsilon}_{L_i \wedge \neg L_j} \nsubseteq \mathcal{N}^{\epsilon'}_{L_i \wedge \neg L_j}$ since $y \in \mathcal{N}^{\epsilon'}_{L_i \wedge \neg L_j}$ and $y \notin \mathcal{N}^{\epsilon}_{L_i \wedge \neg L_j}$. Also, $\mathcal{N}^{\epsilon'}_{L_i \wedge \neg L_j} \nsubseteq \mathcal{N}^{\epsilon}_{L_i \wedge \neg L_j}$ since $x \in \mathcal{N}^{\epsilon}_{L_i \wedge \neg L_j}$ and $x \notin \mathcal{N}^{\epsilon'}_{L_i \wedge \neg L_j}$.

In fuzzy set theory the idea of α-cuts is frequently applied in order to extend classical methods to the fuzzy case [5]. The α-cut of a fuzzy set A, denoted A_α, is given by the set of domain elements with membership in A greater than or equal to α. In label semantics there is an analogous idea where the α-cut of an expression $\theta \in LE$ is given by the set of elements of Ω for which the appropriateness measure of θ is greater than or equal to α i.e. $\theta_\alpha = \{x \in \Omega\colon \mu_\theta(x) \geqslant \alpha\}$.[7] In the following we show that, provided δ satisfies certain smoothness conditions, then there is a direct mapping between the α-cuts θ_α and the neighbourhoods $\mathcal{N}^{\epsilon}_{\theta}$ for expressions $\theta \in LE^{\wedge,\vee}$.

Definition 27. Let $\Delta : [0, \infty) \to [0, 1]$ such that $\Delta(\epsilon) = \delta([\epsilon, \infty))$.

Notice that since Δ is an integral (i.e. $\Delta(\epsilon) = \int_{\epsilon}^{\infty} \delta(\epsilon)\, d\epsilon$) it is a continuous decreasing function into $[0, 1]$ and that for labels $L_i \in LA \mu_{L_i}(x) = \Delta(d(x, P_i))$. This suggests a strong relationship with the threshold model proposed by Hampton [16]

[7] Given the assertability model proposed in Section 3.2, then for $\alpha \geqslant 0.5$ $\theta_\alpha = \{x \in \Omega\colon \mathit{Assert}_\alpha(x \text{ is } \theta)\}$ which corresponds to those elements x for which 'x is θ' can be asserted at certainty level α.

to explain the relationship between the typicality of an instance in a category (as proportional to its similarity to a set of prototypes) and its membership in that category. Specifically, Hampton proposed that the membership of an instance in a category L should be defined as a function m of the similarity of the instance to the prototypes of L, and where m is an increasing function into $[0,1]$. Clearly if we replace similarity with distance from prototypes this would naturally suggest that membership should be a decreasing function m' of distance into $[0,1]$, as is consistent with label semantics prototype model where $\mathsf{m}' = \Delta$. Interestingly, re-analysing an experiment of McCloskey and Glucksberg [25], Hampton [16,17] required subjects to make binary categorization decisions for 482 items into 17 categories. They also independently made assessments of typicality. Category membership was then taken as corresponding to the probability of a positive categorization. In general, the results were consistent with a threshold model where m was a cumulative normal distribution.[8] Notice, also that if δ is assumed to be a normalised normal distribution (as in Example 22) then Δ is one minus the corresponding normalised cumulative normal distribution. Hence, the experimental results of McCloskey and Glucksberg appear to be consistent with individuals making the binary decision that 'x is L_i', for category label L_i, provided that $d(x, P_i) \leqslant \epsilon$ for normally distributed threshold ϵ.

Lemma 28. *$\forall\theta \in LE^{\wedge,\vee}$, $\forall\alpha \in (0,1]$ $(\theta)_\alpha = N_\theta^{\epsilon^*}$ where $\epsilon^* = \sup\{\epsilon\colon \Delta(\epsilon) \geqslant \alpha\}$.*

Proof. Now clearly Δ is differentiable and hence continuous. Therefore, $\{\epsilon\colon \Delta(\epsilon) \geqslant \alpha\} = [0, \epsilon^*]$ where $\epsilon^* = \sup\{\epsilon\colon \Delta(\epsilon) \geqslant \alpha\}$. Hence by Theorems 17 and 20,

$$\begin{aligned}(\theta)_\alpha &= \bigl\{x\colon \mu_\theta(x) \geqslant \alpha\bigr\} = \bigl\{x\colon \delta\bigl(I(\theta,x)\bigr) \geqslant \alpha\bigr\} = \bigl\{x\colon \Delta\bigl(lb(\theta,x)\bigr) \geqslant \alpha\bigr\} \\ &= \bigl\{x\colon lb(\theta,x) \leqslant \epsilon^*\bigr\} = \bigl\{x\colon \epsilon^* \in I(\theta,x)\bigr\} = \bigl\{x\colon \mathcal{D}_x^{\epsilon^*} \in \lambda(\theta)\bigr\} = N_\theta^{\epsilon^*} \qquad \square\end{aligned}$$

Theorem 29. *If δ is such that $\Delta|_J : J \to (0,1]$ is a strictly decreasing function where $\Delta|_J$ is the restriction of Δ to $J = \{\epsilon\colon \Delta(\epsilon) > 0\}$ then $\forall\theta \in LE^{\wedge,\vee}$, $\langle\theta_\alpha, \geqslant, (0,1]\rangle$ is isomorphic to $\langle\mathcal{N}_\theta^\epsilon, \leqslant, J\rangle$.*

Proof. Since $\Delta|_J$ is strictly decreasing and continuous on J then $\Delta|_J : J \to (0,1]$ is a bijection. Consequently, the inverse function $\Delta|_J^{-1} : (0,1] \to J$ is a strictly decreasing bijective function. Hence, for $\epsilon^* = \sup\{\epsilon\colon \Delta|_J(\epsilon) \geqslant \alpha\}$ we have that $\epsilon^* = \Delta|_J^{-1}(\alpha)$. Consequently, by Lemma 28 we have that $\theta_\alpha = \mathcal{N}_\theta^{\Delta|_J^{-1}(\alpha)}$ as required. $\square$

Corollary 30. *For $\Delta|_J$ satisfying the conditions of Theorem 29. If α is a random variable into $(0,1]$ defined such that $\alpha = \Delta|_J(\epsilon)$ then α is a uniformly distributed random variable.*

Proof.

$$\forall\alpha^* \in (0,1] \quad P\bigl(\alpha \leqslant \alpha^*\bigr) = P\bigl(\epsilon \geqslant \Delta|_J^{-1}\bigl(\alpha^*\bigr)\bigr) = \Delta|_J\bigl(\Delta|_J^{-1}\bigl(\alpha^*\bigr)\bigr) = \alpha^* \qquad \square$$

In [5] the use of α-cuts is proposed as a method for extending set functions to fuzzy sets. Given a set function $f : 2^\Omega \to \mathbb{R}$ and a fuzzy set A on Ω the value $f(A)$ is defined by:

$$f(A) = \int_0^1 f(A_\alpha)\, d\alpha$$

This is the expected value of $f(A_\alpha)$ assuming a uniform distribution α. However, no clear justification is given in [5] as to why α should be uniformly distributed. Now the concept of prototype neighbourhoods suggests a method for extending set functions to expressions in *LE* by taking the expected value $f(\mathcal{N}_\theta^\epsilon)$ as follows:

$$f(\theta) = \int_0^\infty f\bigl(\mathcal{N}_\theta^\epsilon\bigr)\delta(\epsilon)\, d\epsilon$$

From Corollary 30 we that, for the restricted class of expressions $LE^{\wedge,\vee}$, this is equivalent to:

$$f(\theta) = \int_0^1 f(\theta_\alpha)\, d\alpha$$

[8] In the case of biological categories, there was also evidence that categorization probability was also affected by non-similarity based knowledge e.g. where an animal looked superficially like one category but was actually in another.

In other words, we see from Corollary 30 that whatever the distribution δ on ϵ, a uniform distribution should be selected when averaging over α-cuts of expressions in $LE^{\wedge,\vee}$.

5.3. Domain information from vague expressions

One of the principle uses of vague concepts in language is to convey information about the underlying universe Ω in specific cases or contexts. Parikh [30] gives the example of two college lecturers Ann and Bob, where Ann asks Bob to bring in her blue book from her library at home. How can Bob use the information that the 'book is blue' to restrict the possible books of interest? The prototype interpretation of label semantics naturally suggests an imprecise restriction on the domain Ω imposed by a constraint 'x is θ' where $\theta \in LE$. Specifically, given the information 'x is θ' then, provided θ is not a contradiction, an agent infers that $x \in \mathcal{N}_\theta^\epsilon$ and consequently that $\mathcal{N}_\theta^\epsilon \neq \emptyset$. From the latter inference the agent naturally generates a posterior distribution δ_θ on ϵ conditional on the information that the neighbourhood of θ is non-empty, where:

$$\forall \epsilon \in [0,\infty) \quad \delta_\theta = \begin{cases} \frac{\delta(\epsilon)}{1-\delta\{\epsilon:\ \mathcal{N}_\theta^\epsilon = \emptyset\}} & \mathcal{N}_\theta^\epsilon \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

Hence, conditioning on the information 'x is θ' results in the following random set model:

Definition 31. For $\theta \in LE$ $\mathcal{N}_\theta{}^\epsilon$ is a random set from the probability space $\langle \mathcal{B}, [0,\infty), \delta_\theta \rangle$ into[9] the probability space $\langle \mathcal{A}_\theta, U_\theta, M_\theta \rangle$ where $\mathcal{B}$ is the σ-algebra of Borel subsets of $[0,\infty)$, δ_θ is defined as above, $U_\theta = \{\mathcal{N}_\theta^\epsilon:\ \epsilon \in [0,\infty)\}$, $\mathcal{A}_\theta = \{\{\mathcal{N}_\theta^\epsilon:\ \epsilon \in I\}:\ I \in \mathcal{B}\}$ and $\forall A \in \mathcal{A}_\theta$ $M_\theta(A) = \delta_\theta(\{\epsilon:\ \mathcal{N}_\theta^\epsilon \in A\})$.

The single point coverage function of the random set in Definition 31 then indicates the probability that a value $x \in \Omega$ is a possible referent in the assertion 'x is θ':

$$\forall x \in \Omega \quad spc_\theta(x) = M_\theta\big(\{\mathcal{N}_\theta^\epsilon:\ x \in \mathcal{N}_\theta^\epsilon\}\big) = \delta_\theta\big(\{\epsilon:\ x \in \mathcal{N}_\theta^\epsilon\}\big) \propto \mu_\theta(x)$$

Notice, from Theorem 26, that in the case of $\theta \in LE^{\wedge,\vee}$ then $\mathcal{N}_\theta^\epsilon$ is nested and therefore $spc_\theta(x)$ is a possibility distribution on Ω. Hence, for this restricted set of expressions our approach is in accordance with that of Zadeh [39] who argues that fuzzy information generates possibilistic constraints on the underlying domain of discourse.

5.4. A general model

Suppose for each label $L_i \in LA$ there is a distinct function $d_i : \Omega^2 \to [0,\infty)$ and a neighbourhood threshold given by random variable ϵ_i. Also suppose $\vec{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$ has a joint density function δ defined on $[0,\infty)^n$. In this model we have a natural definition of the complete description of an element and the neighbourhood of an expression as follows:

Definition 32.

$$\forall x \in \Omega \quad \mathcal{D}_x^{\vec{\epsilon}} = \big\{L_i:\ d_i(x, P_i) \leqslant \epsilon_i\big\} \quad \text{and} \quad m_x(F) = \delta\big(\big\{\vec{\epsilon}:\ \mathcal{D}_x^{\vec{\epsilon}} = F\big\}\big)$$
$$\forall \theta \in LE \quad \mathcal{N}_\theta^{\vec{\epsilon}} = \big\{x \in \Omega:\ \mathcal{D}_x^{\vec{\epsilon}} \in \lambda(\theta)\big\}$$

Notice from Definition 32 that if δ_i denotes the marginal density of ϵ_i derived from joint density δ then the appropriateness measure for label $L_i \in LA$ can be evaluated directly from δ_i according to:

$$\begin{aligned} \mu_{L_i}(x) &= \delta\big(\big\{(\epsilon_1, \ldots, \epsilon_n):\ L_i \in \mathcal{D}_x^{(\epsilon_1,\ldots,\epsilon_n)}\big\}\big) = \delta\big(\big\{(\epsilon_1, \ldots, \epsilon_n):\ d_i(x, L_i) \leqslant \epsilon_i, \epsilon_j \in [0,\infty) \text{ for } j \neq i\big\}\big) \\ &= \delta_i\big(\big[d_i(x, p_i), \infty\big)\big) \end{aligned}$$

The joint density δ encodes dependencies between the random variables ϵ_i which in turn reflect semantic dependencies between the labels. For the consonant model we have total dependence where $d_i = d$ and $\epsilon_i = \epsilon$ for all $L_i \in LA$. This suggests that the labels in LA all describe the same characteristic of the elements in Ω as represented by attribute values in a single conceptual space. Typical examples where such a dependent model may be appropriate are colour labels based on attributes from a multi-dimensional space such as the so-called colour spindle [11] or on rgb values. Other examples include labels for one-dimensional characteristics such height, weight, income and, of course, number of hairs on your head. The following variant on the approach proposed in Section 5.1 also results in a consonant model but where there is variation in scale between the threshold variables ϵ_i.

[9] I.e. for a fixed θ, $\mathcal{N}_\theta^\epsilon$ is a function from $[0,\infty)$ into U_θ which satisfies $\forall I \in \mathcal{B}$, $\{\mathcal{N}_\theta^\epsilon:\ \epsilon \in I\} \in \mathcal{A}_\theta$.

5.4.1. A general consonant model

Assume that $d_i = d$ for all $L_i \in LA$ and that $\epsilon_i = f_i(\epsilon)$ for some increasing function $f_i : [0,\infty) \to [0,\infty)$ of random variable ϵ into $[0,\infty)$. Further suppose that ϵ has density function δ defined on $[0,\infty)$. Then we define:

$$\mathcal{D}_x^\epsilon = \{L_i\colon d(x, P_i) \leqslant f_i(\epsilon)\} \quad \text{so that} \quad m_x(F) = \delta(\{\epsilon\colon \mathcal{D}_x^\epsilon = F\}) \quad \text{and} \quad \mathcal{N}_\theta^\epsilon = \{x \in \Omega\colon \mathcal{D}_x^\epsilon \in \lambda(\theta)\}$$

Now clearly since f_i is an increasing function for $L_i \in LA$ it holds that $\mathcal{D}_x^\epsilon \subseteq \mathcal{D}_x^{\epsilon'}$ for $\epsilon \leqslant \epsilon'$. So $\mathcal{D}_x^\epsilon$ forms a nested hierarchy as ϵ varies.

The following theorem shows how one label L_i can be defined as a restriction of another label L_j in the consonant model so that $\forall \epsilon \geqslant 0$ $\mathcal{N}_{L_i}^\epsilon \subseteq \mathcal{N}_{L_j}^\epsilon$. For example, the label *navy blue* is a restriction of the label *blue* in this sense. In such cases L_j is an appropriate description whenever L_i is appropriate and $\forall x \in \Omega$ $\mu_{\neg L_i \vee L_j}(x) = 1$.

Theorem 33. *For labels $L_i, L_j \in LA$ where $\forall \epsilon \geqslant 0$ $f_j(\epsilon) \geqslant \sup\{d(x, P_j)\colon x \in \mathcal{N}_{L_i}^\epsilon\}$ then $\forall \epsilon \geqslant 0$ $\mathcal{N}_{L_i}^\epsilon \subseteq \mathcal{N}_{L_j}^\epsilon$.*

Proof. For $\epsilon \geqslant 0$ suppose $x \in \mathcal{N}_{L_i}^\epsilon$ then $d(x, P_j) \leqslant f_j(\epsilon) \Rightarrow x \in \mathcal{N}_{L_j}^\epsilon$. $\square$

Notice that one simple case of nested labels is where $P_i \subseteq P_j$ and $f_i \leqslant f_j$. In this case $\sup\{d(x, P_j)\colon d(x, P_i) \leqslant f_i(\epsilon)\} \leqslant \sup\{d(x, P_i)\colon d(x, P_i) \leqslant f_i(\epsilon)\}$ (since $\forall x$ $d(x, P_j) \leqslant d(x, P_j)) \leqslant f_i(\epsilon) \leqslant f_j(\epsilon)$.

5.4.2. An independence model

Suppose for each label $L_i \in LA$ there is a distinct metric $d_i : \Omega^2 \to [0,\infty)$. Further suppose that the threshold distance values for each metric are independent random variables ϵ_i with density δ_i, so that the joint density $\delta = \prod_{L_i \in LA} \delta_i$.

Theorem 34. *If $\epsilon_i\colon L_i \in LA$ are independent random variables then*:

$$\forall F \subseteq LA \quad m_x(F) = \prod_{L_i \in F} \mu_{L_i}(x) \times \prod_{L_i \notin F} (1 - \mu_{L_i}(x))$$

Proof. W.l.o.g. assume $F = \{L_1, \ldots, L_k\}$ then

$$\begin{aligned} m_x(F) &= \delta(\{(\epsilon_1, \ldots, \epsilon_n)\colon \mathcal{D}_x^{(\epsilon_1, \ldots, \epsilon_n)} = \{L_1, \ldots, L_k\}\}) \\ &= \delta(\{(\epsilon_1, \ldots, \epsilon_n)\colon d_1(x, P_1) \leqslant \epsilon_1, \ldots, d_k(x, P_k) \leqslant \epsilon_k, d_i(x, P_i) > \epsilon_i,\ \text{for } i > k\}) \\ &= \prod_{i \leqslant k} \delta_i([d(x, P_i), \infty)) \times \prod_{i > k} \delta_i([0, d(x, P_i))) = \prod_{i \leqslant k} \mu_{L_i}(x) \times \prod_{i > k} (1 - \mu_{L_i}(x)) \qquad \square \end{aligned}$$

Clearly we see from Theorem 34 that the assumption of independence between $\epsilon_i\colon L_i \in LA$ results in a model consistent with the independent selection function (Definition 12). This type of model would seem to be relevant in the case where different labels refer to different independent characteristics of the elements in Ω. For example, if Ω corresponds to a set of people, then judgments concerning the appropriateness of labels *rich* and *tall* would be made on the basis of independent metrics comparing people in terms of their wealth and their height respectively.

5.4.3. Semantic dependence

Semantic dependence between labels as modelled by the joint distribution δ on $\epsilon_i\colon L_i \in LA$ captures the relationship between the various characteristics described by the different labels in *LA*. If all labels describe the same characteristic of elements in Ω, as represented by vectors of attribute values in a shared conceptual space, then an assumption of strong semantic dependence such as in the consonant model or the general consonant model (Section 5.4.1) is appropriate. On the other hand, if each label refers to a different characteristic as represented by points in distinct independent conceptual spaces then an assumption of independence between ϵ_i, as in Section 5.4.2, is valid. For more diverse label sets *LA*, we would expect there to be more complex dependencies between $\epsilon_i\colon L_i \in LA$ than the two extreme cases described above. For example, the characteristics described by different labels may have conceptual spaces which, while not being identical, have attributes in common or which have dependencies between attributes. It is certainly not the case that every joint distribution δ will result in a selection function as given in Definition 10, and hence the resulting calculus for appropriateness measures may not be functional. In such cases it would be interesting to investigate the use of graphical models such as Bayesian Networks to represent the dependencies between the different threshold variables. However, this possibility is beyond the scope this paper and remains to be explored as part of future work.

5.5. Discussion of the label semantics prototype model

Typically in prototype theory the prototypes are not necessarily seen as corresponding to actual examples experienced by the agent but instead as abstractions derived from experience (see Hampton [15]). This is in keeping with many unsupervised learning algorithms where class labels based on clusters are represented by centroids aggregated from actual data. Alternatively, exemplar models (Medin and Shaffer [26]) are a variant on prototype theory where agents have a memory store consisting of actually encountered exemplars of particular description labels. The k-nearest neighbours algorithm is an obvious example of a classification method based on this philosophy. Now the application of prototype theory in this paper does not depend on any distinction between actual exemplars and abstractions provided both are members of the underlying space Ω. Consequently, the label semantics approach can be applied to both variants of prototype theory.

The model presented in Section 5 defines prototypes only for the basic labels LA and not for general compound expressions in $LE - LA$. For expression $\theta \in LE - LA$ the neighbourhood $\mathcal{N}_\theta^\epsilon$ is defined recursively on the basis of distance from prototypes for the labels which occur in θ rather than being based on distance from a set of prototypes for θ itself. This is consistent with the underlying philosophy of label semantics which assumes that the labels LA represent primitive concepts and that the appropriateness of compound expressions is derived as a function only of the appropriateness of these labels. Indeed there is a close analogy here with the distinction that Gärdenfors [11] draws between *properties* which correspond to convex regions of conceptual space, and hence can potentially be represented by prototypes, and *concepts* which are constructed as combinations of properties. However, in its original conception [32] prototype theory assumes that compound expressions such as *not tall* or *orange & not red* are also defined directly in terms of proximity to a (set of) prototypical element(s). In the label semantics framework this would correspond to the assumption that for every expression $\theta \in LE$ there exists prototypes $P_\theta \subseteq \Omega$ such that the set of elements which can be appropriately described by θ are exactly those elements which lie with a threshold ϵ of P_θ i.e. $\mathcal{N}_\theta^\epsilon = \{x\colon\ d(x, P_\theta) \leqslant \epsilon\}$. We shall refer to this assumption as the *compound prototypes hypothesis*.

Upon reflection the compound prototypes hypothesis reveals itself to be problematic in a number of ways: For example, consider negated expressions where $\theta \equiv \neg\varphi$ for some $\varphi \in LE$. In this case, the epistemic stance would require that $\mathcal{N}_{\neg\varphi}^\epsilon$ should not overlap with $\mathcal{N}_\varphi^\epsilon$ (i.e. $\forall\epsilon \geqslant 0$ $\mathcal{N}_\varphi^\epsilon \cap \mathcal{N}_{\neg\varphi}^\epsilon = \emptyset$), since an agent would never be willing to assert both 'x is φ' and 'x is $\neg\varphi$' at any threshold level. This requirement, however, is inconsistent with the compound prototypes hypothesis. To see this consider $y \in P_{\neg\varphi}$ and let $\epsilon \geqslant d(y, P_\varphi)$, then $y \in \mathcal{N}_\varphi^\epsilon$ and $y \in \mathcal{N}_{\neg\varphi}^\epsilon$ and hence $\mathcal{N}_\varphi^\epsilon \cap \mathcal{N}_{\neg\varphi}^\epsilon \supseteq \{y\} \neq \emptyset$. In other words, if $\neg\varphi$ is defined in terms of the distance from a set of prototypes $P_{\neg\varphi}$ then it will always be possible to select a threshold ϵ sufficiently large that $P_{\neg\varphi}$ and $\mathcal{N}_\varphi^\epsilon$ overlap. Another difficulty with the compound prototypes hypothesis arises when considering conjunctions where $\theta \equiv \varphi \wedge \psi$. If we take the view that an agent would only be willing to assert 'x is $\varphi \wedge \psi$' if they are willing to independently assert both 'x is φ' and 'x is ψ' then we should assume that $\forall\epsilon \geqslant 0$ $\mathcal{N}_{\varphi\wedge\psi}^\epsilon \subseteq \mathcal{N}_\varphi^\epsilon \cap \mathcal{N}_\psi^\epsilon$. However, suppose we consider a simple scenario where $\Omega = \mathbb{R}$, $d(x, y) = \|x - y\|$ (Euclidean distance), $P_\varphi = \{a\}$ and $P_\psi = \{b\}$ where $b > a$. We might think of φ and ψ as representing imprecise numbers *about a* and *about b* respectively. In this context it would be rather natural to assume that $P_{\varphi\wedge\psi} = \{c\}$ where $a < c < b$. However, selecting $0 < \epsilon < \frac{d(a,c)}{2}$ we have that $\mathcal{N}_\varphi^\epsilon \cap \mathcal{N}_\psi^\epsilon = \emptyset$ while $\mathcal{N}_{\varphi\wedge\psi}^\epsilon \supseteq \{c\} \neq \emptyset$, so that $\mathcal{N}_{\varphi\wedge\psi}^\epsilon \nsubseteq \mathcal{N}_\varphi^\epsilon \cap \mathcal{N}_\psi^\epsilon$. It should be noted that the requirement $\forall\epsilon \geqslant 0$ $\mathcal{N}_{\varphi\wedge\psi}^\epsilon \subseteq \mathcal{N}_\varphi^\epsilon \cap \mathcal{N}_\psi^\epsilon$ is only valid since $\varphi \wedge \psi$ is a true conjunctive. In natural language it is often the case that compound expressions which would appear to be conjunctions from a syntactic perspective, do not semantically correspond to classical conjunctions. For example, an expressions such as *small elephant*, while appearing to be in the form of a conjunction, clearly does not refer to elements of a domain which can be both independently referred to as being *small* and as being an *elephant*. Here the label *small* is in some sense secondary to the label *elephant*, whereby the latter identifies the context in which the former is defined. In a recent paper Freund [10] refers to this ordering of properties in the definition of a concept as *determination*. This notion would also seem to be relevant to the type of expressions studied in Hampton [18], where experimental results suggest that descriptions such as *office furniture* are not treated as classical conjunctions. A detailed treatment of such expressions is, however, beyond the scope of this paper and the question of how to extend label semantics so as to include them remains to be explored in future work.

Adopting the compound prototypes hypothesis also requires us to consider how an agent could determine the prototypes P_θ for every expression $\theta \in LE$. In the label semantics model the implicit assumption is that for the basic labels $L_i \in LA$, P_i would be determined by a learning process involving interaction and communication with other agents. Using compound prototypes could potentially provide an agent with more flexible and relevant concept definitions if they could base their choice of prototypes on actual experience of language use. The difficulty here, however, is the potentially huge amounts of data required to determine P_θ for every semantically distinct expression $\theta \in LE$. For instance, if $|LA| = n$ then there 2^{2^n} semantically distinct expressions in LE. For an agent to obtain specific data on the use of each of these would seem to be infeasible even for moderately large n. Consequently, it would seem likely that to some extent agents would need to define P_θ for compound expression θ based on the prototypes of its component expressions. In effect this would require some variant of the functionality assumption that for each $\theta \in LE$ there is a prototype generating function $g_\theta : 2^\Omega \times \cdots \times 2^\Omega \to 2^\Omega$ according to which $P_\theta = g_\theta(P_i\colon\ L_i \in LA)$. A special case of functionality would then be full prototype compositionality whereby generating functions $g_\wedge : 2^\Omega \times 2^\Omega \to 2^\Omega$, $g_\vee : 2^\Omega \times 2^\Omega \to 2^\Omega$ and $g_\neg : 2^\Omega \to 2^\Omega$ would be defined, according to which $P_{\theta\wedge\varphi} = g_\wedge(P_\theta, P_\varphi)$, $P_{\theta\vee\varphi} = g_\vee(P_\theta, P_\varphi)$ and $P_{\neg\theta} = g_\neg(P_\theta)$. Now aside from the difficulty of defining these generating functions in an intuitive manner, as discussed at length in Lawry [22], it is difficult to see what advantages as a representation framework a functional approach to prototype generation has over the kind of recursive definition of neighbourhoods described in Section 5.

6. Conclusions

We have argued that concept vagueness in natural language is a manifestation of uncertainty about the appropriate use of labels, which naturally arises as a result of the distributed and example based manner in which language is learnt. For this reason we adopt an epistemic viewpoint according to which the uncertainty associated with vague concepts is quantified by a measure of belief that the relevant labels or expressions are appropriate to describe a given example, and where appropriateness is governed by the emergent conventions of language use. The label semantics framework has then been introduced to provide a formal calculus for appropriateness measures of this kind.

Prototype theory is an effective tool by which we can understand the role of similarity and typicality in the definition of natural categories. The use of uncertain thresholds on the distance between elements and prototypes in order to define boundaries for labels provides a clear link between prototype theory and the epistemic view of vagueness, as well as providing an intuitive interpretation of appropriateness measures. This prototype model can be understood from two perspectives characterised by the random sets $\mathcal{D}_x^\epsilon$ and $\mathcal{N}_\theta^\epsilon$ respectively. For an element $x \in \Omega$, $\mathcal{D}_x^\epsilon$ identifies those labels in LA with prototypes sufficiently close (i.e. within ϵ) to x for them to be deemed appropriate descriptions of x. On the other hand, for a given expression θ the neighbourhood $\mathcal{N}_\theta^\epsilon$ contains those elements for which θ is an appropriate description. Both approaches provide characterisations of appropriateness measures as single point coverage functions of random sets. Intuitively, an agent would focus on $\mathcal{D}_x$ when attempting to identify appropriate descriptions of a specific instance that they are presented with. Alternatively, the agent would focus on $\mathcal{N}_\theta^\epsilon$ when making inferences about the underlying domain Ω, given an assertion 'x is θ'.

By allowing different distance functions and thresholds for each label in *LA* we can model the semantic dependence between labels by a joint distribution on the cross product space of threshold random variables. In this paper we have discussed in detail two specific types of semantic dependence each of which results in a particular selection function for mass values. These are *total dependence* where all threshold variables and distance metrics coincide and *total independence* where all threshold variable are statistically independent. The semantic dependence being modelled here is based on the relationship between the characteristics being described by the various labels, as represented by different conceptual spaces. So for the consonant (dependent) model all labels describe the same characteristic, while for the independent model all labels describe different completely independent characteristics. Intermediate cases of dependence may not result in a selection function and consequently the resulting calculus for appropriateness measures will not be functional.

As already mentioned in Section 5.4.3 one of the most interesting and promising areas to be explored as part of future studies are the richer models of semantic dependence allowed within the general framework proposed in Section 5.4. The use of graphical approaches such as Bayesian networks to model the joint distribution on threshold variables, while not necessarily resulting in a functional calculus, may provide computationally tractable methods for evaluating appropriateness measures. Another potential avenue for research is the development of rule-based methods incorporating the label semantics prototype theory. In this case new rule-learning methods could be investigated which combine clustering algorithms with label semantics to provide the definition of natural description labels.

Acknowledgements

We would like to thank the reviewers for their many insightful comments on earlier drafts of this paper.

Yongchuan Tang is funded by the National Natural Science Foundation of China (NSFC) under Grant 60604034, the joint funding of NSFC and MSRA under Grant 60776798, and the Science and Technology Program of Zhejiang Province under Grant 2007C223061.

References

[1] B. Berlin, P. Kay, Basic Color Terms: Their Universality and Evolution, University of California Press, 1977.
[2] G. Coletti, R. Scozzafava, Conditional probability, fuzzy sets, and possibility: A unifying view, Fuzzy Sets and Systems 144 (2004) 227–249.
[3] B. de Finetti, Fondamenti logici del ragionamento probabilistico, Bollettino dell' Unione Matematica Italiana 9 (1930) 258–261.
[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, 1988, pp. 742–755.
[5] D. Dubois, H. Prade, Measuring properties of fuzzy sets: A general technique and its use in fuzzy query evaluation, Fuzzy Sets and Systems 38 (1990) 137–152.
[6] D. Dubois, H. Prade, The three semantics of fuzzy sets, Fuzzy Sets and Systems 90 (2) (1997) 141–150.
[7] D. Dubois, L. Godo, H. Prade, F. Esteva, An information-based discussion of vagueness, in: H. Cohen, C. Lefebre (Eds.), Handbook of Categorization in Cognitive Science, 2005, pp. 892–913.
[8] K. Fine, Vagueness, truth and logic, Synthese 30 (1975) 265–300.
[9] T. Flaminio, L. Godo, A logic for reasoning about the probability of fuzzy events, Fuzzy Sets and Systems 158 (6) (2007) 625–638.
[10] M. Freund, On the notion of concept I, Artificial Intelligence 172 (2008) 570–590.
[11] P. Gärdenfors, Conceptual Spaces: The Geometry of Thought, MIT Press, 2000.
[12] I.R. Goodman, H.T. Nguyen, Uncertainty Model for Knowledge Based Systems, North Holland, 1985.
[13] I.R. Goodman, Fuzzy sets as equivalence classes of random sets, in: R. Yager (Ed.), Fuzzy Set and Possibility Theory, 1982, pp. 327–342.

[14] P. Hájek, Ten questions and one problem on fuzzy logic, Annals of Pure and Applied Logic 96 (1999) 157–165.
[15] J.A. Hampton, Concepts as prototypes, in: B.H. Ross (Ed.), The Psychology of Learning and Motivation: Advances in Research and Theory, vol. 46, 2006, pp. 79–113.
[16] J.A. Hampton, Typicality, graded membership and vagueness, Cognitive Science 31 (3) (2007) 355–384.
[17] J.A. Hampton, Similarity-based categorization and fuzziness of natural categories, Cognition 65 (1998) 137–165.
[18] J.A. Hampton, Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion, Journal of Experimental Psychology: Learning, Memory and Cognitio 14 (1) (1988) 12–32.
[19] E. Hisdal, Are grades of membership probabilities, Fuzzy Sets and Systems 25 (1988) 325–348.
[20] R. Kintz, J. Parker, R. Boynton, Information transmission in spectral color naming, Perception and Psychophysics 5 (1969) 241–245.
[21] J. Lawry, A framework for linguistic modelling, Artificial Intelligence 155 (2004) 1–39.
[22] J. Lawry, Modelling and Reasoning with Vague Concepts, Springer, 2006.
[23] J. Lawry, Appropriateness measures: An uncertainty model for vague concepts, Synthese 161 (2) (2008) 255–269.
[24] J. Lawry, An overview of computing with words using label semantics, in: H. Bustince, F. Herrera, J. Montero (Eds.), Fuzzy Sets and Their Extensions: Representation, Aggregation and Models, Springer, 2008, pp. 65–87.
[25] M. McCloskey, S. Glucksberg, Natural categories: Well defined or fuzzy sets? Memory and Cognition 6 (1978) 462–472.
[26] D.L. Medin, M.M. Schaffer, Concept theory of classification learning, Psychological Review 85 (3) (1978) 207–238.
[27] H.T. Nguyen, On modeling of linguistic information using random sets, Information Science 34 (1984) 265–274.
[28] V. Novák, Towards formalized integrated theory of fuzzy logic, in: Z. Bien, K. Min (Eds.), Fuzzy Logic and its Applications to Engineering, Information Sciences, and Intelligent Systems, Kluwer, Dordrecht, 1995, pp. 353–363.
[29] R. Parikh, Vague predicates and language games, Theoria (Spain) XI (27) (1996) 97–107.
[30] R. Parikh, Vagueness and utility: The semantics of common nouns, Linguistics and Philosophy 17 (1994) 521–535.
[31] F.P. Ramsey, in: R.B. Braithwaite (Ed.), The Foundations of Mathematics, and Other Logical Essays, Kegan Paul, Trench, Trubner and Company Ltd., London, 1931.
[32] E.H. Rosch, Natural categories, Cognitive Psychology 4 (1973) 328–350.
[33] E.H. Rosch, Cognitive representation of semantic categories, Journal of Experimental Psychology: General 104 (1975) 192–233.
[34] E.H. Ruspini, On the semantics of fuzzy logic, International Journal of Approximate Reasoning 5 (1991) 45–88.
[35] L. Sivik, C. Taft, Color naming: A mapping in the NCS of common color terms, Scandinavian Journal of Psychology 35 (1994) 144–164.
[36] Y. Tang, J. Zheng, Linguistic modelling based on semantic similarity relation amongst linguistic labels, Fuzzy Sets and Systems 157 (2006) 1662–1673.
[37] T. Williamson, Vagueness, Routledge, 1994.
[38] L.A. Zadeh, Fuzzy sets, Information and Control 8 (3) (1965) 338–353.
[39] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1978) 3–28.
[40] L.A. Zadeh, The concept of linguistic variable and its application to approximate reasoning, Part 1, Information Sciences 8 (1975) 199–249.
[41] L.A. Zadeh, The concept of linguistic variable and its application to approximate reasoning, Part 2, Information Sciences 8 (1975) 301–357.
[42] L.A. Zadeh, The concept of linguistic variable and its application to approximate reasoning, Part 3, Information Sciences 9 (1976) 43–80.