# Possible Worlds and Possible Meanings:

**Brandon Bennett**
School of Computing
University of Leeds
brandon@comp.leeds.ac.uk

**Abstract**

The paper develops a formal model for interpreting vague languages based on a variant of *supervaluation* semantics. Two modes of semantic variability are modelled, corresponding to different aspects of vagueness: one mode arises where there can be multiple definitions of a term; and the other relates to the threshold of applicability of a vague term with respect to the magnitude of relevant observable values.

The truth of a proposition depends on both the possible world and the *precisification* with respect to which it is evaluated. Structures representing possible worlds and precisifications are both specified in terms of primitive functions representing observable measurements, so that the semantics is grounded upon an underlying theory of physical reality. On the basis of this semantics, the acceptability of a proposition to an agent is characterised in terms of a combination of agent's beliefs about the world and their attitude to admissible interpretations of vague predicates.

## 1 Introduction

The terminology of natural language is highly affected by vagueness. Except in specialised circumstances, there are generally no agreed criteria that precisely determine the applicability of our conceptual vocabulary to describing the world. This presents a considerable problem for the construction of a knowledge representation language that is intended to articulate information of a similar kind to that conveyed by natural language communication. Hence, in so far as many kinds of AI system require an interface with some natural form of human communication, the processing of vague information is a central problem for AI.

The fundamental idea of the *supervaluationist* account of vagueness, is that a language containing vague predicates can be interpreted in many different ways, each of which can be modelled in terms of a precise version of the language, which is referred to as a *precisification*. If a classical semantics is used to give a denotational valuation of expressions for each of these precise versions, the interpretation of the vague language itself is given by a *supervaluation*, which is determined by the collection of these classical valuations.

Copyright © 2011, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

The view that vagueness can be analysed in terms of multiple senses was proposed by Mehlberg (1958), and a formal semantics based on a multiplicity of classical interpretations was used by van Fraassen (1969) to explain 'the logic of presupposition'. It was subsequently applied to the analysis of vagueness by Fine (1975), and thereafter has been one of the more popular approaches to the semantics of vagueness adopted by philosophers and logicians. In the current paper we introduce a logic based on the essential idea of supervaluation semantics, but different in several respects from previous systems.

A major strength of the supervaluation approach is that it enables the expressive and inferential power of classical logic to be retained (albeit within the context somewhat more elaborate semantics) despite the presence of vagueness. In particular, necessary logical relationships among vague concepts can be specified using classical axioms and definitions. These analytic interdependencies will be preserved, even though the criteria of correspondence between concepts and the world are ill-defined and fluid.

Investigation of supervaluation semantics in the philosophical literature tends, as one might expect, to be drawn towards subtle foundational questions (such as those concerning the *sorites* paradox and second-order vagueness). By contrast, the purpose of the current paper is to flesh out the details of a particular variant of supervaluation semantics and to develop an expressive formal representation language that could be employed within information processing applications.

The development of the supervaluation idea in the current paper also departs somewhat from that proposed by Fine. In Fine's theory precisifications vary in their level of precision, so that one precisification may be a more precise version of another. This gives rise to a partial order on precisifications. Fine then proposes a semantics that takes account of this ordering and defines a notion of *super-truth* in terms of the precisification structure as a whole: super-truth corresponds to truth at all maximally precise and *admissible* precisifications (where 'admissible' means that a precisification is considered a reasonable interpretation of the language and is taken as a primitive notion). Moreover, Fine suggests that 'truth' in a vague language may be identified with this notion of super-truth.

By contrast, in the current paper, we take each precisifi-

cation to be a maximally precise version of the language. And we consider truth primarily as a property of propositions that is relative to a particular precisification, rather than determined by the whole set of possible precisifications. However, we will also introduce the notion of a proposition holding relative to a *standpoint*, which is associated with a set of precisifications considered acceptable by some agent. This notion is formally somewhat akin to 'super-truth', except that admissibility is understood as relative to a particular agent in a particular situation.

Like supervaluation semantics, standpoint semantics may be regarded as a rival to the *fuzzy logic* approach to semantic indeterminacy (Zadeh 1975). Whereas fuzzy logic explores this indeterminacy in terms of degrees of truth and non-classical truth functions, standpoint semantics focuses on truth conditions rather than truth values, and employs a notion of truth which is close to the classical view, although relativised to account for a variety of possible interpretations. Nevertheless, we believe that strong correspondences between standpoint semantics and fuzzy logic can be established. By introducing a probability distribution over the space of precisifications, degrees of *acceptability* can be introduced, and it can be shown that the acceptability of vague conjunctions is governed by certain of the $T$-norms commonly used in fuzzy logic. Results in this area are beyond the scope of the present work, where we focus on the core model theory and employ a much simpler model of an agent's attitude to precisifications.

The formalism developed in the current paper takes ideas from previous theories proposed by Bennett (1998; 2006) and Halpern (2004). Halpern's paper analyses vagueness in terms of the subjective reports of multiple agents, but these play a similar role in his semantics to precisifications in the semantics proposed in this paper. Our approach also has some commonality with that of Lawry (2008) and Lawry & Tang (2009).

In (Bennett 2006) the semantics of vague adjectives is characterised in terms of their dependence on *relevant objective observables* (e.g. 'tall' is dependent on 'height'). One of the primary aims of the current paper is to provide a rigorous foundation for the notion of precisification, in which the interpretation associated with a precisification is explicitly defined in terms of choices made in imposing distinctions with regard to continuously varying properties manifest in possible states of the world. Bennett (1998) proposed a two-dimensional model theory in which the interpretations of propositions are indexed both by precisifications and possible worlds. Whereas a somewhat *ad hoc* relation of relevance between between vague predicates and observables was introduced in (Bennett 2006), the current paper makes a much more specific connection, in which thresholds occur explicitly in definitions of vague predicates. A concrete example of the use of this approach in an implemented computer system for processing geographic information can be found in (Santos, Bennett, & Sakellariou 2005) and (Bennett, Mallenby, & Third 2008).

The structure of the paper is as follows: in the next section we give an overview of the formal theory that will be developed, and consider some examples illustrating different kinds of vagueness. In Section 3 we specify a formal language that makes explicit the structure of both possible worlds and precisifications in terms of possible values of observable measurements. Section 4 gives a formal model of an agent's *standpoint* with respect to possible worlds that the agent considers plausible and precisifications that the agent considers admissible. We end with a consideration of further work and conclusions.

## 2 Preliminaries

Before getting into the details of our formal language and its semantics, we first clarify some aspects of our approach.

**Comparison Classes** To avoid confusion we briefly consider a phenomenon that is often associated with vagueness but will not be considered in the current paper. This is the relativity of the interpretation of vague adjectives to a given *comparison class*. When describing a woman as tall we employ a different threshold of tallness than when describing a giraffe. As in this example, the relevant comparison class is often determined by the count noun used to refer to the object, but it may sometimes be determined by the particular set of objects present in a given situation. However, comparison class relativity is a side-issue that is not essential to vagueness itself. Even if we restrict attention to a definite class of individuals (say, adult males in Belgium) the adjective 'tall' is still vague. Similar remarks could be made about more general issues of context variability of the interpretation of terminology. If required, an explicit model of comparison class dependency of vague adjectives (perhaps similar to that given in (Bennett 2006)) could be incorporated into an extended version of our theory.

**Distinguishing Conceptual and Sorites Vagueness** An important feature of the proposed theory is that it makes a clear distinction between two forms of vagueness.

One type of vagueness arises where there is ambiguity with regard to which attributes or conditions are essential to the meaning of a given term, so that it is controversial how it should be defined. We call this *conceptual vagueness* (or 'deep ambiguity'). A good example of conceptual vagueness is the concept of *murder*. Although in most cases there will be general agreement as to whether a given act constitutes murder, the precise definition is subtle and controversial. Judicial systems vary as to the stipulations they make to characterise the crime of murder. Thus one may debate whether murder requires malice or intent, whether the murderer must be sane, whether the victim must be unwilling e.t.c.. Moreover, even where conditions are stipulated in great detail, cases may arise that defy simple judgement.

A somewhat different kind of vagueness occurs when the criteria for applicability of a term depend on placing a threshold on the *required magnitude* of one or more variable attributes. For instance, we may agree that the appropriateness of ascribing the predicate 'tall' to an individual depends on the height of that individual, but there is no definite height threshold that determines when the predicate is applicable. We refer to this as *sorites vagueness*, since the essence of the sorites paradox is the indeterminacy in the number of grains required to make a heap.

So to summarise: in the case of *conceptual vagueness* there is indeterminism regarding which property or logical combination of properties is relevant to determining whether a concept is applicable, whereas with *sorites vagueness* the relevant properties are clear, but the degree to which these properties must be present is indefinite.[1]

Of course, the two modes of vagueness are closely related. From a theoretical point of view, one could regard the setting of thresholds as just a part of the definition of a term, and thus just a special form of conceptual vagueness. Furthermore, many terms are affected by both kinds of vagueness. For example, Kamp (1975) considers the conditions under which one might describe a person as 'clever'. Here it is not clear what parameter or parameters are relevant to the attribution of cleverness. As Kamp suggests, quick-wittedness and problem solving ability are both indications of cleverness, although one person might be considered more quick-witted that another and yet less capable of problem solving. Moreover, even if we were to decide upon which of these or which combination of these (and other) attributes were required for cleverness, we would still need to set a threshold on the degree to which these attributes must be present.

Despite their close connection, there are significant differences in the type of semantic variability involved — to reiterate: conceptual vagueness is indeterminacy in the attribute or combination of attributes that must be present, whereas sorites vagueness is indeterminacy in the degree to which a continuously varying attribute (or attributes) must be present. Hence, we believe that a semantics for vague languages is most clearly specified by separating the two modes.

**Predication, Observables and Thresholds** Our semantics explicitly models the applicability of vague predicates in terms of thresholds applied to relevant observable measurements. In the simplest case we assume that our judgement of whether a predicate $\phi$ applies to object $x$ depends only on the value of an measurement $f(x)$ — the higher the value of $f(x)$, the more we are inclined to judge that $\phi(x)$ is true. Let $\tau(\phi)$ denote some reasonable threshold that we might set for the applicability of $\phi$. Then $\phi(x)$ is judged to be true if $f(x)$ is greater than $\tau(\phi)$ and false if $f(x)$ is less than $\tau(\phi)$.

The case where we have $f(x) = \tau(\phi)$ presents a technical issue, in that there is no obvious basis to decide between assigning truth or falsity. This will be avoided by restricting the semantics so that a threshold value may not have the same value as any observable function.

**Relating Precisifications to Cognitive Attitudes** As well as relating precisifications to states of the world we also model their relationship to the cognitive states of agents. We give an account of an agent's attitude to vague propositions in terms of a formalised notion of *standpoint*, which describes the agents belief state as well as the range of interpretations of vague terminology that they consider admissible. A *standpoint* will be modelled by a structure $\langle B, A, \theta \rangle$, where: $B$ is the set of possible worlds compatible with the agent's beliefs; $A$ is the set of precisifications that are acceptable to the agent; and $\theta$ is a semantic theory encoding axioms and definitions that govern the agent's language use.

As is usual in supervaluation-based approaches, we assume that when describing a particular situation or appraising a given set of propositions, a language user employs a choice of thresholds that is consistent across usages of all concepts. Thus, where two or more concepts have some semantic inter-dependence, this will be maintained by consistent usage of thresholds. For example 'tall' and 'short' are (in a particular context) mutually exclusive and are dependent on the objective observable of height. Thus the height threshold above which a person to be considered 'tall' must be greater than the height threshold below which a person is considered 'short'. However, we do not assume that an agent's point of view is determined by a single precisification but rather by a set of accepted thresholds (or, as in our further work, by a probability distribution over the set of all precisifications).

## 3 A Language of Precise Observables and Vague Predicates

In this section we define a general-purpose formal language that (despite bearing only a coarse correspondence to the structure and meaning of natural language) is intended to exhibit some fundamental principles that govern the phenomenon of vagueness.

A key idea underlying the construction of this formalism is that the language should contain two types of vocabulary:

- A precise vocabulary for describing the results of precise objective measurements of the state of the world.
- A vague vocabulary which is defined in terms of the precise vocabulary, relative to a valuation of certain *threshold parameters*, which may occur in the definitions.

### Measurement Structures

At the base of the semantics is a structure that represents the state of a *possible world* in terms of a valuation of *measurement functions*, which specify the results of observations applied to the entities of some domain.

An *$n$-ary measurement function* over domain $D$ is a function $\mu : D^n \to \mathbb{Q}$, with $\mathbb{Q}$ being the set of rational numbers. Thus, $\mathbb{Q}^{D^n}$ is the set of all $n$-ary measurement functions and $\bigcup_{n \in \mathbb{N}} \mathbb{Q}^{D^n}$ is the set of all measurement functions of any arity (with domain $D$).

A *measurement structure* is a tuple $\mathcal{M} = \langle D, M, \nu_M, w \rangle$, where:

- $D$ is a domain of entities;
- $M = \{\ldots, f_i, \ldots\}$ is a set of measurement function symbols;
- $\nu_M : M \to \mathbb{N}$, is a mapping from the symbols in $M$ to the natural numbers, giving the arity of each function;
- $w : M \to \bigcup_{n \in \mathbb{N}} \mathbb{Q}^{D^n}$, such that if $\nu_M(f) = m$ then $w(f) \in \mathbb{Q}^{D^m}$, is a function mapping each $n$-ary function symbol to a measurement function from $D^n$ to $\mathbb{Q}$.

[1] This distinction was identified and analysed in (Bennett 2005), but no formal semantics was presented.

Since each assignment function $w$ characterises the domain and function symbols, as well as determining a valuation of the measurement functions over the domain, we regard each $w$ as representing a *possible world*.[2] Given a domain $D$, a set of measurement function symbols $M$, and an arity specification function $\nu_M$, the set of worlds that can be specified in terms of these elements can be defined by[3]

$$\mathsf{Worlds}(D, M, \nu_M) = \{w \mid \langle D, M, \nu_M, w\rangle \text{ is a measurement structure}\} .$$

A *measurement frame* is a structure that specifies all possible worlds determined by a given measurement structure:

$$\langle D, M, \nu_M, W\rangle ,$$

where $W = \mathsf{Worlds}(D, M, \nu_M)$.

## A Language of Measurements and Thresholds

Let us now consider the definition of a predicative language that can be interpreted relative to a measurement structure.

Let $\mathcal{L}(M, \nu_M, T, V)$ be the set of formulae of a first-order logical language[4] whose non-logical symbols consist of: a finite set of measurement function symbols $M = \{f_1, \ldots, f_k\}$, a finite set, $T = \{\mathsf{t}_1, \ldots \mathsf{t}_l\}$, of *threshold parameter* symbols, strict and non-strict inequality relations ($<$ and $\leq$), and a denumerable set $V = \{\ldots, x_i, \ldots\}$ of variable symbols. Every atomic formula of $\mathcal{L}(M, \nu_M, T, V)$ takes one of the forms:

**A1.** $f_j(x_1, \ldots, x_n) \leq f_k(y_1, \ldots, x_m)$

**A2.** $\mathsf{t}_i \leq \mathsf{t}_j$

**A3.** $\mathsf{t}_i < f_j(x_1, \ldots, x_n)$

**A4.** $f_j(x_1, \ldots, x_n) < \mathsf{t}_i$

where $n = \nu_M(f_j)$ and $m = \nu_M(f_k)$.

Nested functions are not allowed, since measurement functions operate on entities and their values are real numbers. Complex formulae are formed from atomic formulae by means of the standard truth-functional connectives and quantifiers over the variables (but not over the threshold parameters). $\mathcal{L}(M, \nu_M, T, V)$ includes formulae with free variables.

Although the full language $\mathcal{L}(M, \nu_M, T, V)$ allows arbitrary logical combinations of the different kinds of atoms, certain restrictions on the atoms give rise to distinct types of formulae. Those containing only atoms of the form **A1** can be regarded as expressing constraints on the physical structure of the world. Such formulae (especially those where all variables are universally quantified) can be used to express a *physical theory*.[5] [6] Formulae of the form **A2** express ordering constraints between thresholds. A typical example would be $\mathsf{t_short} \leq \mathsf{t_tall}$, stating that the threshold below which an individual is considered short is less than the threshold above which and individual is considered tall. A set of such formulae will be called a *threshold constraint theory*.

## Predicate Definitions

Each formula of $\mathcal{L}(M, \nu_M, T, V)$ defines a predicate of arity $n$, where $n$ is the number of free variables in the formula. Hence, we can extend $L$ by defining new predicate symbols by means of formulae of the form

**PG.** $\forall x_1, \ldots, x_n[R(x_1, \ldots, x_n) \leftrightarrow \Phi(\mathsf{t}_1, \ldots, \mathsf{t}_m, x_1, \ldots, x_n)]$ ,

where $\Phi(\mathsf{t}_1, \ldots, \mathsf{t}_m, x_1, \ldots, x_n)]$ is any formula in $\mathcal{L}(M, \nu_M, T, V)$ incorporating parameters $\mathsf{t}_1, \ldots, \mathsf{t}_m$ and with free variables $x_1, \ldots, x_n$.

For instance, $\forall x[\mathsf{Tall}(x) \leftrightarrow (\mathsf{t_tall} \leq \mathbf{height}(x))]$ is a typical example of a predicate defined in this way. Here, $\mathbf{height}$ is a measurement function and $\mathsf{t_tall}$ is a threshold parameter. An informal interpretation of this formula is that an entity is tall just in case its height is greater than or equal to the value of the parameter $\mathsf{t_tall}$.

## The Full Language Including Defined Predicates

Let $\mathcal{L}(M, \nu_M, T, V, R, \nu_R, N)$ be the language obtained by supplementing the vocabulary of $\mathcal{L}(M, \nu_M, T, V)$ by a set of predicate symbols $R$, such that the arity of each symbol is given by the function $\nu_R : R \to \mathbb{N}$ and a set of constant symbols $N$. The set of atomic formulae is extended to include those of the form

**A5.** $R_i(\alpha_1, \ldots, \alpha_n)$,

where each $\alpha_i \in (V \cup N)$. The complete set of formulae of $\mathcal{L}(M, \nu_M, T, V, R, \nu_R, N)$ includes all formulae constructed from this extended set of atomic formulae by application of Boolean connectives and quantification over variables in $V$.

**Predicate Grounding Theories** Given a language $\mathcal{L}(M, \nu_M, T, V, R, \nu_R, N)$, a *predicate grounding theory* for this language is a set of formulae of the form **PG**, containing one formula for each relation $R_i \in R$. Thus, the predicate grounding theory defines every relation in $R$ in terms of a formula of the sub-language $\mathcal{L}(M, \nu_M, T, V)$.

Let $\Theta$ be the set of all predicate grounding theories for $\mathcal{L}(M, \nu_M, T, V, R, \nu_R, N)$. Since each of these grounding theories includes a definition of every predicate in the language, we can define a function,

$$\mathsf{Def} : \Theta \times R \to \mathcal{L}(M, \nu_M, T, V) ,$$

[2] Note that a possible world in this sense is an *arbitrary* valuation of the function symbols over the domain. The valuation need not respect physical laws with regard to possible combinations of measurable values, so, in so far as the observable functions are intended to correspond to actual kinds of measurements, such worlds could be physically impossible.

[3] Here I am making the simplifying (unrealistic) assumption that we have the same set of entities present in every possible world.

[4] In fact, this account does not depend on the specific details the language. I choose standard first-order logic for definiteness, but there could be reasons to use a more expressive language.

[5] A fully fledged physical theory would require additional mathematical operators, but this complication will not concern us here.

[6] Here we are making the assumption that all measurements are comparable. A more realistic theory would model the fact that only certain measurements can be usefully compared. We could restrict the form **A1** to the case where the values of functions $f_i$ and $f_j$ must relate to the same units of measurement.

such that $\mathsf{Def}(\theta, R)$ is a formula with $\nu_R(R)$ free variables, which gives the definition of the relation $R$.

## Parameterised Precisification Models

We now define a model structure to provide a semantics for $\mathcal{L}(M, \nu_M, T, V, R, \nu_R, N)$. The model incorporates a measurement frame together with mappings from the language symbols onto elements of the frame. Specifically, a *parameterised precisification model* is a structure

$$\mathfrak{M} = \langle \mathcal{M}, R, \nu_R, N, V, T, \Theta, \kappa, \xi, P \rangle \ ,$$

where

- $\mathcal{M} = \langle D, M, \nu_M, W \rangle$ is a measurement frame;
- $R$ is a set of predicate symbols;
- $\nu_R : R \to \mathbb{N}$ gives the arity of each predicate symbol;
- $N$ is a set $\{\ldots, \mathsf{n}_i, \ldots\}$ of nominal constants;
- $V$ is a set $\{\ldots, x_i, \ldots\}$ of variable symbols;
- $T$ is a finite set $\{\ldots, \mathsf{t}_i, \ldots\}$ of threshold parameter symbols;
- $\Theta = \{\ldots, \theta_i, \ldots\}$, where each $\theta_i$ is a predicate grounding theory for the language;
- $\kappa : N \to D$ maps nominal constants to entities of the domain;
- $\xi : V \to D$ maps variable symbols to entities of the domain;
- $P = \{p \mid p : T \to (\mathbb{R} \backslash \mathbb{Q})\}$, is the set of all mappings from threshold parameters to *irrational* numbers. ($P$ is the set of precisifications.)

The assignment of irrational numbers to threshold parameters is a technical means to ensure that every observable value is either greater or smaller than any threshold parameter. For practical applications it would probably be better to use two separate sub-domains of the rationals, but this would require a lengthier exposition.

## Interpretation Function

The semantic interpretation function, $[\![\chi]\!]^{w,p,\theta}_{\mathfrak{M}}$, gives the denotation of any formula or term $\chi$ of the language relative to a given model $\mathfrak{M}$, a possible world $w \in W$, a predicate grounding theory $\theta \in \Theta$ and a precisification $p \in P$. The $\theta$ index models semantic indeterminacy arising from conceptual vagueness, whereas the $p$ index models indeterminacy due to sorites vagueness.

To specify the interpretation function, the following auxiliary notations will be used:

- $\mathfrak{M} \overset{x}{\sim} \mathfrak{M}'$ means that models $\mathfrak{M}$ and $\mathfrak{M}'$ are identical except for their variable assignment functions $\xi$ and $\xi'$. And moreover, these assignment functions are identical, except that they may differ in the value assigned to the variable $x$.
- $\mathsf{Subst}([x_1 \Rightarrow \alpha_1, \ldots, x_n \Rightarrow \alpha_n], \phi)$ refers to the formula resulting from $\phi$ after replacing each variable $x_i$ by $\alpha_i$.

The interpretation function can now be specified. The Boolean connectives and quantifier have their standard classical interpretation:

- $[\![\neg\phi]\!]^{w,p,\theta}_{\mathfrak{M}} = \mathbf{t}$ if $[\![\phi]\!]^{w,p,\theta}_{\mathfrak{M}} = \mathbf{f}$, otherwise $= \mathbf{f}$ ;
- $[\![\phi \wedge \psi]\!]^{w,p,\theta}_{\mathfrak{M}} = \mathbf{t}$ if $[\![\phi]\!]^{w,p,\theta}_{\mathfrak{M}} = \mathbf{t}$ and $[\![\psi]\!]^{w,p,\theta}_{\mathfrak{M}} = \mathbf{t}$, otherwise $= \mathbf{f}$ ;
- $[\![\forall x[\psi]]\!]^{w,p,\theta}_{\mathfrak{M}} = \mathbf{t}$ if $[\![\psi]\!]^{w,p,\theta}_{\mathfrak{M}'}$ for all $\mathfrak{M}'$ such that $\mathfrak{M} \overset{x}{\sim} \mathfrak{M}'$, otherwise $= \mathbf{f}$.

The inequality relations are interpreted as follows, where $\gamma_i$ and $\gamma_j$ may be either a threshold parameter or a measurement function term:

- $[\![\gamma_i \leq \gamma_j]\!]^{w,p,\theta}_{\mathfrak{M}} = \mathbf{t}$ if $[\![\gamma_i]\!]^{w,p,\theta}_{\mathfrak{M}}$ is less than or equal to $[\![\gamma_j]\!]^{w,p,\theta}_{\mathfrak{M}}$, otherwise $= \mathbf{f}$ ;
- $[\![\gamma_i < \gamma_j]\!]^{w,p,\theta}_{\mathfrak{M}} = \mathbf{t}$ if $[\![\gamma_i]\!]^{w,p,\theta}_{\mathfrak{M}}$ is strictly less than $[\![\gamma_j]\!]^{w,p,\theta}_{\mathfrak{M}}$, otherwise $= \mathbf{f}$.

The value of measurement functions depends on the possible world in which the measurement is made; and hence, their interpretation depends on the $w$ index:

- $[\![f(\alpha_1, \ldots, \alpha_n)]\!]^{w,p,\theta}_{\mathfrak{M}} = w(f)(\langle \delta(\alpha_1), \ldots, \delta(\alpha_n) \rangle)$, where $\delta(\alpha) = \kappa(\alpha)$ if $\alpha \in N$ and $\delta(\alpha) = \xi(\alpha)$ if $\alpha \in V$.

Interpretation of the threshold parameters depends on the precisification index, $p$:

- $[\![\mathsf{t}]\!]^{w,p,\theta}_{\mathfrak{M}} = p(\mathsf{t})$ ;

Finally, the interpretation of the defined predicate and relation symbols is dependent upon the grounding theory $\theta$:

- $[\![R(\alpha_1, \ldots, \alpha_n)]\!]^{w,p,\theta}_{\mathfrak{M}} = [\![\mathsf{Subst}([x_1 \Rightarrow \alpha_1, \ldots, x_n \Rightarrow \alpha_n], \mathsf{Def}(\theta, R))]\!]^{w,p,\theta}_{\mathfrak{M}}$

On the basis of the interpretation function, a semantic satisfaction relation can be defined by

$$\mathfrak{M}, \langle w, p, \theta \rangle \Vdash \phi \quad \text{iff} \quad [\![\phi]\!]^{w,p,\theta}_{\mathfrak{M}} = \mathbf{t} \ .$$

This says that formula $\phi$ is true in model $\mathfrak{M}$, at world $w$ and precisification $p$, with predicate grounding theory $\theta$.

The *interpretation set* of a proposition relative to a model $\mathfrak{M}$ is given by:

$$[\![\phi]\!]_{\mathfrak{M}} = \{\langle w, p, \theta \rangle \mid (\mathfrak{M}, \langle w, p, \theta \rangle \Vdash \phi)\}$$

This is the set of world/precisification/grounding theory triples for which formula $\phi$ is evaluated as true.

We have now established the main result of this paper: we have defined a first-order language with a semantics that gives a special status to observable measurements and threshold parameters. The interpretation function for this language is such that each valuation of observable measurements corresponds to a possible world, and each valuation of threshold parameters corresponds to a precisification. In terms of this semantics each propositional formula is interpreted as the set of possible world/precisification/grounding theory triples at which the proposition is considered to be true.

## 4 Standpoints and Proposition Evaluation

In order to take account of an agent's comprehension of propositional information, we need to relate the agent's cognitive state to our formal semantics of propositions, which gives the meaning of a proposition in terms of an interpretation set. Two aspects of the cognitive state are clearly relevant: what the agent believes about the state of the world, and what the agent regards as an acceptable usage of terminology (especially vague predicates). Whether the agent considers a proposition to be true will depend on both these aspects.

Beliefs may be modelled either syntactically in terms of formulae expressing facts and theories that an agent regards as true, or semantically in terms of possible states of the world that an agent considers plausible. In the framework of classical logic, where each predicate has a definite meaning, the two perspectives are tightly linked, since any set of formulae determines a fixed set of possible worlds that satisfy that set. But if propositions can vary in meaning, according to different interpretations of vague predicates, the correspondence is more fluid: different interpretations will be true in different sets of possible worlds. Thus, in order to separate an agent's beliefs about the world from their attitude to linguistic meanings, the beliefs must be modelled in a way that is not affected by linguistic variability. Hence, our model of belief is primarily based on sets of plausible possible worlds rather than theories. Of course the structure of a possible world will still be determined relative to a formal language, but this will be the limited language of observable measurement functions, which contains no predicates other than precise ordering relations.

In accordance with the interpretation function, $[\![\chi]\!]_{\mathfrak{M}}^{w,p,\theta}$, specified above, the agent's attitude to the meanings of vocabulary terms is modelled in terms of both the predicate grounding definitions and the choices of threshold values that the agent considers to be acceptable.

We represent an agent's attitude by a structure that we call a *standpoint* which comprises a tuple,

$$\langle B, A, \Psi \rangle \ ,$$

where:

- $B \subseteq W$ is the agent's *belief set* — i.e. the set of possible worlds that are compatible with the agent's beliefs,[7]
- $A \subseteq P$ is the agent's *admissibility set* — i.e. the set of precisifications that the agent considers to make reasonable assignments to all threshold parameters, and hence to be *admissible*,
- $\Psi \subseteq \Theta$ is a set of predicate grounding theories that characterises all possible definitions of ambiguous predicates that the agent regards as acceptable.

Thus, a standpoint characterises the range of possible worlds and linguistic interpretations that are plausible/acceptable to the agent.

[7]Here we use a very simple model of belief. In further development it might be useful to introduce a richer belief theory. We may want to distinguish an agents implicit beliefs, from explicit propositional belief (along the lines of (Fagin & Halpern 1988).

For any given model $\mathfrak{M}$, we can now formally define the condition that a formula $\phi$ holds with respect to a particular standpoint $\langle B, A, \Psi \rangle$. Specifically, we define:

- $\mathfrak{M}, \langle B, A, \Psi \rangle \Vdash \phi$ iff $(B \times A \times \Psi) \subseteq [\![\phi]\!]_{\mathfrak{M}}$.

So $\phi$ holds for a standpoint if it is true at all worlds in the belief set for all admissible precisifications and all acceptable predicate grounding theories. In other words the agent considers that for any reasonable interpretation of ambiguous predicates and all choices of threshold parameters, $\phi$ is true in all possible worlds consistent with the agent's beliefs.

**Weaker forms of Assertion Relative to a Standpoint** Our standpoint semantics also enables us to specify a number of modal-like operators by means of which we can describe more ambivalent and/or less confident attitudes that an agent may have to a given proposition:

- $\mathfrak{M}, \langle B, A, \Psi \rangle \Vdash \mathsf{CouldSay}(\phi)$ iff $(B \times \{p\} \times \{\theta\}) \subseteq [\![\phi]\!]_{\mathfrak{M}}$, for some $p \in A$ and some $\theta \in \Psi$.
- $\mathfrak{M}, \langle B, A, \Psi \rangle \Vdash \mathsf{CouldBe}(\phi)$ iff $(\{w\} \times A \times \Psi) \subseteq [\![\phi]\!]_{\mathfrak{M}}$, for some $w \in B$.
- $\mathfrak{M}, \langle B, A, \Psi \rangle \Vdash \mathsf{CouldBeSay}(\phi)$ iff $\langle w, p, \theta \rangle \in [\![\phi]\!]_{\mathfrak{M}}$, for some $w \in B$, $p \in A$ and $\theta \in \Psi$.

$\mathsf{CouldSay}(\phi)$ asserts that for all worlds in the agent's belief set, $\phi$ is true in some admissible precisification for some acceptable grounding theory. This operator is used to characterise an assertion made in a context where an agent is fully confident that their beliefs relevant to $\phi$ are correct, but is unsure about the choice of words used to express $\phi$. By contrast, $\mathsf{CouldBe}(\phi)$ means that, for all reasonable interpretations of predicate definitions and thresholds, there is some world compatible with the agent's beliefs where $\phi$ is true. In this case the interpretation of the words used to express $\phi$ is taken to be uncontroversial, but the state of reality, which would determine whether $\phi$ is true, is uncertain. Finally, $\mathsf{CouldBeSay}(\phi)$ indicates that there is some combination of acceptable predicate definition and threshold choice and some world state compatible with the agent's beliefs according to which $\phi$ would be true.

This distinction between the operators CouldBe and CouldSay is closely related to distinctions made by J.L. Austin [1953], in his analysis of different ways in which the sense of a predicate may be related to the properties of an object to which it is applied. He introduced the idea of the *onus of match* between the sense of a word and the corresponding property of the object, and suggested that in some speech situations one is clear about the meaning of the word but unsure whether the object possesses the appropriate property, whereas in others one is clear about the properties of the object but unsure about whether the word adequately describes the object.

A number of other modalities could be defined. In the specifications for $\mathsf{CouldSay}(\phi)$ and $\mathsf{CouldBeSay}(\phi)$ the indices giving the precisification $p$ and grounding theory $\theta$ are both allowed to vary (independently). But we could, for instance, define a modality $\mathsf{M}$ such that $\mathsf{M}(\phi)$ is true iff $(B \times A \times \{\theta\}) \subseteq [\![\phi]\!]_{\mathfrak{M}}$ for some particular $\theta \in \Psi$ —

i.e. there is some acceptable grounding theory, relative to which $\phi$ holds for every admissible precisification. Such a modality does not have an obvious informal interpretation, since in ordinary language we tend to conflate conceptual and sorites within the general phenomenon of vagueness; however, it may still provide an informative characterisation of an agent's attitude to a proposition.

## 5 Further Work and Conclusions

The theory presented so far provides a core semantic framework within which various significant aspects of vagueness can be articulated. As such, it is intended to provide a platform from which more practical knowledge representation languages can be developed. Such development would most likely involve both simplification of some parts of the semantics and elaboration of others.

An obvious extension is to add probability distributions over the belief set and/or the set of admissible precisifications to model the relative plausibility of different possible worlds and the relative acceptability of different precisifications. Indeed, such an extension is the subject of ongoing work and initial investigations seem to be very fruitful and also to be necessary in order to model the famous *sorites paradox* and related phenomena.

Another interesting direction for further work would be to study the assimilation of new information in terms of the transformation from an agent's initial standpoint to a modified standpoint. Here the issue arises that when an agent receives information that is incompatible with their standpoint, they must choose whether to modify their beliefs or to try to interpret the information from the point of view of a different standpoint. This study could extend to more general aspects of the exchange of information between two agents in the presence of vagueness. The notion of *context* is likely to be relevant to such an investigation.

For practical applications it may be convenient to replace the somewhat elaborate model theory we have given with a more standard first-order semantics. This would require that the semantic indices (worlds and precisifications) were in some way incorporated into the object language. This could be achieved in a similar way to how temporal logics are often treated within AI formalisms (e.g. by a Situation Calculus style formulation such as $\mathsf{Holds}(\phi, \langle w, p, \theta \rangle)$ ).

In summary, the main contributions of this paper have been: development of a formal semantics for interpretation of vague languages that models both the definitional ambiguity of conceptual terms and the variability of thresholds on their applicability; formalisation of an explicit link between these thresholds governing predicate applicability and observable properties of the world; a model of the cognitive standpoint of an intelligent agent incorporating both a belief state and an attitude towards the interpretation of vague terms.

## Acknowledgements

Partial support of the *Co-Friend* project ( FP7-ICT-214975, `www.cofriend.net`) and the EU Framework 7 is gratefully acknowledged.

## References

Austin, J. L. 1953. How to talk: some simple ways. *Proceedings of the Aristotelian Society*. Also in J.L. Austin Philosophical Papers, Clarendon Press, 1961.

Bennett, B.; Mallenby, D.; and Third, A. 2008. An ontology for grounding vague geographic terms. In Eschenbach, C., and Gruninger, M., eds., *Proceedings of the 5th International Conference on Formal Ontology in Information Systems (FOIS-08)*. IOS Press.

Bennett, B. 1998. Modal semantics for knowledge bases dealing with vague concepts. In Cohn, A. G.; Schubert, L.; and Shapiro, S., eds., *Principles of Knowledge Representation and Reasoning: Proceedings of the 6th International Conference (KR-98)*, 234–244. Morgan Kaufman.

Bennett, B. 2005. Modes of concept definition and varieties of vagueness. *Applied Ontology* 1(1):17–26.

Bennett, B. 2006. A theory of vague adjectives grounded in relevant observables. In Doherty, P.; Mylopoulos, J.; and Welty, C. A., eds., *Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning*, 36–45. AAAI Press.

Fagin, R., and Halpern, J. Y. 1988. Belief, awareness, and limited reasoning. *Artificial Intelligence* 34:39–76.

Fine, K. 1975. Vagueness, truth and logic. *Synthèse* 30:263–300.

Halpern, J. Y. 2004. Intransitivity and vagueness. In *Principles of Knowledge Representation: Proceedings of the Ninth International Conference (KR-2004)*, 121–129.

Kamp, H. 1975. Two theories about adjectives. In Keenan, E., ed., *Formal Semantics of Natural Language*. Cambridge, England: Cambridge University Press.

Keefe, R., and Smith, P. 1996. *Vagueness: a Reader*. MIT Press.

Lawry, J., and Tang, Y. 2009. Uncertainty modelling for vague concepts: A prototype theory approach. *Artificial Intelligence* 173(18):1539–1558.

Lawry, J. 2008. Appropriateness measures: an uncertainty model for vague concepts. *Synthese* 161:255–269.

Mehlberg, H. 1958. *The Reach of Science*. University of Toronto Press. Extract on Truth and Vagueness, pp. 427-55, reprinted in (Keefe & Smith 1996).

Santos, P.; Bennett, B.; and Sakellariou, G. 2005. Supervaluation semantics for an inland water feature ontology. In Kaelbling, L. P., and Saffiotti, A., eds., *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, 564–569. Edinburgh: Professional Book Center.

van Fraassen, B. C. 1969. Presupposition, supervaluations and free logic. In Lambert, K., ed., *The Logical Way of Doing Things*. New Haven: Yale University Press. chapter 4, 67–91.

Zadeh, L. A. 1975. Fuzzy logic and approximate reasoning. *Synthese* 30:407–428.