Towards a Probabilistic Semantics for Vague Adjectives

Peter R. Sutton*

King's College London, Department of Philosophy,
Strand, London WC2R 2LS, United Kingdom

Abstract. A way of modelling the meanings of vague terms directly in terms of the uncertainties they give rise to is explored. These uncertainties are modelled with a probabilistic, Bayesian version of situation semantics on which meaning is captured as the correlations between uses of words and the types of situations those words are used to refer to. It is argued that doing so provides a framework from which vagueness arises naturally. It is also claimed that such a framework is in a position to better capture the boundarylessness of vague concepts.

Keywords: Vagueness, Uncertainty, Probabilistic Semantics.

1 Introduction

A seemingly characteristic feature of vague terms is their *boundarylessness* [1]. However, a challenge for truth-conditional accounts of the semantics for vague terms is that truth-conditions create sharp boundaries. One way to try to assuage this tension is to introduce uncertainty into semantics. For example, one could maintain a threshold for truth, but model vagueness as uncertainty about where the threshold, in some context, lies [2] [3]. However, arguably, uncertainty over thresholds nonetheless carves boundaries. In this paper, I will avoid thresholds and develop a way of introducing uncertainty directly into semantic representations. On this approach, sentences are taken to encode uncertainty over types of situations. This can been seen from two perspectives. Either as uncertainty about how the world is, given the way words are used, or as uncertainty about how to use words when faced with a type of situation to describe. The information that words convey is understood as a reflection of the correlations there are between features or properties in the world, and our uses of words to describe or refer to objects with those properties. By adopting this view, I will suggest that

* I would like to thank the organisers and participants of the Bayesian Natural Language Semantics and Pragmatics workshop held at ESSLLI 2013 and the anonymous reviewers for the workshop and this volume for the helpful comments and improvements that they have suggested. Thanks also to the participants of the King's College London Language and Cognition Seminar, with special thanks to Alex Clark, Ruth Kempson, Shalom Lappin, and Wilfried Meyer-Viol for their invaluable comments on earlier versions of this paper.

we can better capture the boundarylessness of terms, and that borderline cases, another characteristic feature of vagueness, emerge directly from the semantics.

In §§2-3 the uncertainty associated with vague terms will be described. In §§4-5, these informal ideas will be modelled using a probabilistic version of situation semantics. In §6, the approach will be linked to vagueness. In §7, the proposal will be compared with two related proposals in the recent literature.

2 Uncertainty

There are two ways that uses of a vague term can give rise to uncertainty. The first concerns the world and the second concerns the word. Say that we are told that John is tall, or that Mary's car is green. Given that we are competent in English, and, given that we have no reason to expect what we have been told to be incorrect, what we have been told makes it reasonable to believe certain things about the world. In the vagueness literature, when the semantics of vague terms are analysed, some less vague properties, concepts, or features are often cited in the analysis. For example, with respect to 'green', our judgements vary over the shades that things are; for 'tall', over the heights that things are; for 'bald', over how much (head) hair individuals have. On being told that John is tall, or that Mary's car is green, we are uncertain about what feature or property John has with respect to height, or that Mary's car has with respect to shade.

This uncertainty could be thought of as a series of graded beliefs. In the following, I will represent the content of these beliefs with reference to measures that may not have any psychological reality. For example, I will talk about someone's beliefs about heights measured in centimetres. However, it need not be assumed that our doxastic representations of heights refer to centimetres. If I talk of Mary's beliefs about John's height as distributing over heights in centimetres, this is merely a convenient notation. It does not require that Mary can say, in centimetres, how tall John is (or might be).

It is, however, possible that using such precise values introduces an artificial level of precision. It could be that the way we cognitively represent the world is also vague, in which case, there should be a mapping from vague words to a mapping from a vague representational level to properties in the world (such as being some height in centimetres or being some particular shade). However, by describing mappings between vague words and less or non-vague properties in the world directly, I am, at worst, oversimplifying the matter by skipping over a mental/cognitive level of representation. For at least some cases we ought to be able to describe how the information conveyed by words relates to the (more or less) precise properties to be found in situations in the world.

It would be unreasonable to believe that John is a specific height just on the basis of being told that he is tall, and it would be unreasonable to believe Mary's car to be a specific shade just on the basis of being told it is green. Herein lies the first kind of uncertainty:

(U1) Descriptions using vague expressions leave us uncertain about specific features/properties that objects in the world have.

U1 uncertainty is often, to some extent, eliminable in practice. We can, for example, go and look at John (maybe even measure him), or we can go and have a look at Mary's car. And yet, even if we are in command of these facts, uncertainty about how to use our language can still remain. If John is 180cm tall, or if Mary's car is an odd sort of turquoisey shade, we might feel thoroughly uncertain whether or not 'tall' would be an effective word to use to describe John or 'green' an effective word to use to describe Mary's car.

By using the term 'effective', I mean to appeal to a notion of success. What is effective (for some purpose) is what succeeds in doing something/bringing about the desired result. What is effective for describing John, or Mary's car is going to be an interest relative matter, since what is effective for doing something in one situation may not be effective in another. A lot of the vagueness literature, insofar as it predominantly concerns solutions to the sorites paradox, takes the only important criterion for what is effective to be what description would be true. However, truth is not always what we need to establish, or are interested in. For example, if our aim is to communicate which individual John is in a crowd, irrespective of whatever story about the truth-conditions for 'tall' are, one might be able to use 'tall' to identify John because he is significantly taller than those around him. Here, what makes a description count as true will not be addressed in detail. Instead, I will continue to use the broader notion of *effective descriptions* which can be used to describe a second kind of uncertainty:

(U2) Specific features/properties that objects in the world have can leave us uncertain about how to effectively describe them.

Relations hold between U1 and U2:

1. Were we equally certain of applying 'tall' to 185cm John as to 190cm Bill, then, all else being equal, we should be equally certain that John (Bill) is 185cm in height when described as tall as 190cm, and *vice versa*
2. Given 1, where U1 similarities give rise to U2 similarities and *vice versa*, we should expect U1 differences to give rise to U2 differences and *vice versa*.
3. Variations in U1 uncertainty can give rise to U2 uncertainty. Being told that John is tall may leave us uncertain about his height. We might be fairly certain that he is not around 170cm, comparatively certain that he is around 190cm, but highly uncertain about whether he is around 180cm. Given 1 and 2, this creates U2 uncertainty in the use of 'tall'. If we are fairly certain that 'tall' effectively describes someone who is around 190cm in height, but fairly certain that 'tall' wouldn't effectively describe someone around 170cm in height, then we should uncertain that 'tall' would effectively describe someone around 180cm. Furthermore, we should also expect a range of U2 judgements to give rise to U1 uncertainty.

As emphasised by an anonymous reviewer, the presence of U1 uncertainty is not sufficient for vagueness, and the presence of U2 uncertainty is not necessary for vagueness. The relationship between vagueness and different kinds of uncertainty will be made clear in §6. In brief, however, there will be times when, given the way the uses of words correlate with types of situations, nothing in the meanings of the words themselves, or any outside information, will be able to resolve our uncertainty. For example, for some heights of people, we may have irresolvable uncertainty about whether to call them tall. This can arise because when someone is described as tall, we may be no more certain that they are that height than we would be were they to be described as not tall.

Two challenges arise for communicating with vague expressions. As a hearer, we must face uncertainty over what objects in the world are like, given the way they have been described. As speakers, we may, from time to time, face uncertainty over how to effectively describe things as being. However, arguably this uncertainty stems directly from the uses of such terms, and from how such uses correlate with the properties of situations those words describe or refer to. In the following, I will treat the strengths of these correlations *as* the information that those terms carry.[1] This information will be modelled using Bayesian conditional probabilities. It is in this sense that Bayesian notions will be at the heart of the semantics for vague terms.

3 Constituents

A semantics based on uncertainty will have to be able to attribute, to types of expression, a role in the larger constructions in which they occur. Take our above examples: 'John is tall' and 'Mary's car is green'. Giving a semantics for the modifiers 'tall' and 'green' can then be taken to be a matter of accounting for what information they carry about some object. For example, 'x is green' might be modelled as making it reasonable to believe that x is one of some rough range of shades.

However, here a disparity between 'tall' and 'green' arises. The x in 'x is tall' is vital for getting any idea about what it is reasonable to believe. If we are only told that something, anything, is tall, be it a mountain, a molehill, a mouse, or a millipede, there is no height it would be more reasonable to believe this thing is than others.[2] This changes as soon as 'tall' is predicated of an NP or is applied to a CN in an NP: The 'tall' in 'x is a tall man' seems to make it more reasonable to believe that x is some heights rather than others. But this suggests that part of the information which contributes to our expectations of heights is coming from the CN/predicate 'man'. This can be seen by substitution of CNs/predicates.[3] Compare how expectations of heights differ for 'x is a tall man', 'x is a tall

[1] This notion of semantic information is more rigorously developed in [4].

[2] This may be to strong. There could be some priors we have, or reasoning we could engage in, that would make some heights more reasonable than others. My thanks go to Noah Goodman for helpful discussion.

[3] I do not mean to take, as assumed, a clear, semantically important division between CNs and adjectives. Below, I will treat words like 'tall' and 'green' as predicate modifiers and words like 'car' and 'man' as as predicates. In many cases, the traditional classifications of words as CNs and adjectives will overlap with these semantic divisions. Nothing I say rests on whether they all do.

molehill' and 'x is a tall mountain'. The situation for green seems to be similar. CNs make a difference to expectations.[4] However, 'x is green' does seem to carry some expectations of its own: that x is roughly within some range of shades.

I take this to be a reason to see some differences in the semantics of various vague adjectives/modifiers. The modifier 'green', which seems to carry information on its own, will have a slightly different semantic shape to those modifiers (such as 'tall') that only seem to modify expectations based on information carried by nominal predicates. The details of this difference will be elaborated in §5.

What we have here is an argument for treating some modifiers as, in some sense, not really carrying information *per se*. Instead we can see them as encoding a modification on the information carried by a nominal predicate. Someone's being described as a woman gives us a rough expectation as to her height. This may be a very broad range, but we have some expectations nonetheless. Something's being described as a skyscraper gives us very different expectations as to its height.

One might worry that this puts a lot of weight on information carried by nominal predicates and/or background information and beliefs. However, all that is being assumed is that learning to categorise and classify objects with such predicates, in part, amounts to developing expectations as to basic visual cues such as size, shape and shade. An anonymous reviewer rightly points out that this approach begins to blur the boundaries between what counts as meaning and what counts as general knowledge about the world. A firmer distinction can be established again without overly affecting the account, however. Rather that seeing the meaning of, say, 'man' as being all the information one has learnt about men, we can view meanings as procedures (see [6]) for accessing such knowledge and for determining the ways such knowledge should be combined.

The common modifiers 'old', 'big' and 'long', just like 'tall', only seem to give us reasonable expectations when applied to a nominal predicate (or when themselves predicated of an NP in cases such as 'John is tall'). In turn, however, that suggests that nominal predicates such as 'man' carry information about what features men can reasonably be expected to have. Some tests (albeit ones based on linguistic intuitions) can be applied to get a grasp on what information this is. For example, at least for non-metaphorical uses, 'x is a long man' does not seem to make much sense/provide us with reasonable expectations about x's length.[5] This suggests that 'man' does not carry information pertaining to length. This is not to say that modifiers such as 'green' are not restricted either. For example, synaesthesia aside, 'x has green ideas' is hard to understand if 'green' relates to shade.[6]

---

[4] This point is related to the commonly held idea that vague adjectives are context-sensitive or interpretable only relative to a comparison class [5].

[5] In English at least, length is not simply a euphemism for height. Trains, passageways, halls, to name a few, can have a great length without having a great height.

[6] An interesting possibility is that the application of domain specific modifiers to NPs that do not carry information about that domain will generate metaphorical

Nominal predicates will be treated as encoding information based on correlations between things that those common nouns are used to classify and the features that the things they are used to classify have. For example, it is because there is a stronger correlation between men and some heights than there is with others, that 'man' carries information about heights. Adjectives will be modelled as modulating some specific aspect of the information that nominal predicates carry. For example, 'tall' will modify information relating to height carried by predicates like 'man'.

A possible worry can be flagged here.[7] In the literature on adjectives, it is common to appeal to a comparison class to capture their semantics [5], [8], [9], [10]. However, it has been pointed out that common nouns do not always determine comparison classes [11]. For example, BMWs are not the comparison class below, despite featuring as part of a modified NP:

1. Kyle's car is an expensive BMW, though it's not expensive for a BMW. In fact it's the least expensive model they make. [11]

The worry concerns whether the above analysis conflates modified nouns with comparison classes. The analysis can, however be modified slightly. Rather than demand of some adjectives that they simply modify a nominal predicate, we can instead require that there is some class of things to which it is being applied in any context (information about which can be modified by the adjective). This still leaves a difference with terms like 'red' which seem to carry some information (about, say, shades) independently of a lexically provided nominal predicate or a comparison class.[8]

4 Semantics: Preliminaries

4.1 Correlations and Situations

In the following, I will suggest one way to formalise U1 uncertainty (uncertainty over how the world is, given a description of it). I will borrow heavily from situation semantics [12], [13], [14], but will incorporate a probabilistic element into the standard determinate theory.

In situation theory, meaning is captured via the notion of constraints. Constraints represent information channels, for example, types of situations in which there is rain in the city are informationally connected to types of situation in which the pavements are wet. The (context independent) linguistic meaning of

interpretations. Metaphor, humour and other such subjects are outside of the scope of this paper. There may also be an interesting link to be explored between the ideas put forward here and work done on scalar adjectives in [7].

[7] This was pointed out by an anonymous reviewer for the BNLSP 2013 Workshop.

[8] That is not to say that 'red' cannot relate to non-stereotypical shades (such as in 'red onion'). Elsewhere [4], I suggest a way for this account to model the information carried by 'pet fish' where goldfish are not the most stereotypical fish, nor the most stereotypical pets.

an expression is, held to be just a special case of a normal information channel: types of situation in which some expression is uttered are informationally related to types of situations in which some conditions obtain. For example, types of situations in which someone utters 'It's raining in London' are informationally connected to types of situations in which there is rain in (at least part of) London (at some time).

It is via this type-type link or relation that agents are able to extract token-token information. For the case in hand, if one is in a token discourse situation of the type *'It's raining in London' is uttered*, one will be led to expect a token situation of the type *raining in London*. In standard situation semantics, ignoring a lot of details, the truth of what is said turns on whether the token described situation is of the right type.

The basic idea to be presented in this paper is that meaning can be treated as correlational. Instead of constraints holding between one utterance situation type and one described situation type, probability theory will be employed to capture different strengths of connections between an utterance situation type, and many described situation types. The meanings of expressions are therefore held to be correlations between discourse situations of a certain type, and described situations of a certain type. For example, the type of situation in which 'John is tall' is uttered will be correlated, to different extents, with types of situation in which John is one of many possible heights.

In Situation Theory, situations support *infons*. Infons are traditionally conceived as what contribute to forming informational items [14], or as types of situations [13].[9] Another way to view them is as properties of situations (as opposed to properties of individuals). For example, (ignoring times) one property that a situation might have is where it's raining somewhere:[10]

$$\langle\!\langle \mathtt{rain}, \dot{l}, yes \rangle\!\rangle$$

The $\dot{l}$ is a *location parameter*. Parameters can be seen as akin to variables. They can be bound via type abstraction (see below), or free. Ignoring quantifiers, free parameters are those not bound by abstraction in situation theoretic propositions. If a situation, $\mathtt{s}$ has some property (some infon), σ, then it is said that the situation supports the infon. That a situation supports an infon, in notation, is a situation theoretic proposition:

$$\mathtt{s} \models \sigma$$

For the above infon, this would give:

$$\mathtt{s} \models \langle\!\langle \mathtt{rain}, \dot{l}, yes \rangle\!\rangle$$

[9] This way of viewing infons (as types) propagates through into situation theoretic approaches with richer type systems. See, for example [15].

[10] This notation for infons is essentially Devlins. However, for polarities, I adopt Barwise and Perry's 'yes' and 'no' instead of Devlin's '1', '0'. This is to avoid potential confusion with the limit cases of probability values $[0, 1]$.

Which says that in situation $\mathtt{s}$, it is raining at some location. Situations are meant to be ways to classify objects and events in the world. They are not possible worlds. If a situation, in fact, has the property σ (is, in fact, of type σ/supports σ), then the proposition is true. However, we will not be seeking to associate linguistic expressions with propositions, nor will we be directly interested in particular situations and whether they support some infon. This is because constraints are defined between situation types.

Importantly, abstraction in situation theory allows one to talk about types. Abstracting over objects (understood loosely as individuals, locations, and times), creates *object types*. Taking our rain example, and abstracting over $\dot{l}$, this gives:

$$\lambda[\dot{l}](\mathtt{s} \vDash \langle\langle \mathtt{rain}, \dot{l}, yes \rangle\rangle)$$

Which is the type of locations in which it rains in situation $\mathtt{s}$. When bound, parameters can be replaced with constants as with standard β-reduction. For example:

$$\begin{aligned} &\lambda[\dot{l}](\mathtt{s} \vDash \langle\langle \mathtt{rain}, \dot{l}, yes \rangle\rangle)\ .\ [\mathtt{London}] \\ \Rightarrow \quad &\mathtt{s} \vDash \langle\langle \mathtt{rain}, \mathtt{London}, yes \rangle\rangle \end{aligned}$$

Situations can be abstracted over to give *situation types*[11] via *situation type abstraction*. Notation varies, I will adopt the following:[12]

$$\lambda[\dot{s}](\dot{s} \vDash \sigma)$$

This means the type of situation in which σ obtains. The $\dot{s}$ is a parameter, as $\dot{l}$ above, but $\dot{s}$ is a parameter for situations. In this case, $\dot{s}$ is bound. For example:

$$\lambda[\dot{s}](\dot{s} \vDash \langle\langle \mathtt{rain}, \mathtt{London}, yes \rangle\rangle)$$

Which is the type of situation in which it rains in London. Correlations (information channels) hold between types. For example the above type might correlate with the type of situation in which the streets of London are wet. In the probabilistic version of Situation Theory to be developed, these correlations will be modelled as conditional probabilities (the probability of one type, given another). It is this that will ensure that if some concrete situation of a type arises, there will be a probability that some situation of another type will arise (or, arose, or will have arisen etc.).

The proposal to be made here is then that U1 uncertainty (where descriptions using vague expressions leave us uncertain about specific features/properties that

[11] More recent situation theoretic approaches take types to be objects. An example of this rich type-theoretic approach is Type Theory with Records (TTR) [15].

[12] This is close to Devlin's notation, but with Cooper's use of 'λ' for abstractions instead of Devlin's $(\dot{s}|\dot{s} \vDash \sigma)$. This variation is to avoid possible confusion resulting from the use of '|' in probability theory.

objects in the world have), can be modelled as conditional probabilities between types of situation. I will say more on U2 probability in §7.

We will have situations of two kinds:

(i) *Discourse situations*, $\mathtt{d}$, are situations in which some specific discourse takes place. For example, we might have the discourse situation in which 'red' is used to classify some object, or, in which 'tall' is used to describe some person. Discourse situations will model the situations in which certain words are used (certain things are said).

(ii) *Described situations*, $\mathtt{s}$, are situations in which the world is some way. For example, we might have the described situation in which an object is some shade of colour, or, in which a person is 180cm in height. Described situations will model the kinds of ways the world can be.

Types can be formed by abstracting over discourse and described situations. For example:

$$\lambda[\dot{d}](\dot{d} \vDash \langle\langle \text{utters}, \dot{a}, \text{TALL}, \mathtt{j}, yes \rangle\rangle)$$

Is the type of situation in which someone describes John as 'tall'. Given such a type, there will be a probability of some described situation being of some type. For example, the type of described situation in which John is 180cm in height:

$$\lambda[\dot{s}](\dot{s} \vDash \langle\langle \mathtt{height{=}180cm}, \mathtt{j}, yes \rangle\rangle)$$

Varying probabilities will result from different height values. These probabilities will form a distribution. Abstracting away from the contribution made by the name 'John', this distribution will reflect the extent to which uses of the modifier 'tall' to describe, say, a human male, correlates with human males being certain heights. The heights that an agent will entertain will be restricted by their learning process (namely, an approximation over the kinds of humans, males etc. that they have experienced). It will be supposed that a rational agent can only have distributions that sum to 1.

The information carried by sentences will be captured as the conditional probability of there being a described situation of some type, given a discourse situation of some type. For example, the probability of there being a situation type in which John is 5ft, given a situation type in which John is described as tall, will be (comparatively) very low.[13] Other conditional probabilities will be greater for greater height properties John might have. A range of such probabilities (each with a different height property/one-place relation assigned to John in the described situation) will form a distribution (must sum to one).

Conditional probability values will represent something like credence. However, rather than just a degree of belief, we will be concerned with the degree to

[13] This can be so even given the assumption that speakers are not being deliberately deceptive. What is being tracked here is the extent to which, in general, properties such as heights of individuals correlate with types of utterances, such as describing them as tall.

which it is reasonable to believe that some state of affairs obtains. This distinction is between what, given some actual description, a hearer believes (credence), and what, in virtue of the information carried by the utterance of the words, one is entitled to believe. Conditional probability values reflect correlations between uses of words and states of affairs. This incorporates all uses, both representations and misrepresentations of the world. The degree to which it is reasonable that some state of affairs obtains reflects this. How much reason or entitlement one has to believe the world to be some way will be a direct reflection of how words are used (how uses of words correlate with types of situations).

So, if being told that John is tall makes it reasonable to believe to a degree of n that he is 180cm in height, this will be represented as assigning a probability value of n to the conditional probability of there being a situation of the type in which John is 180cm in height, given a situation of the type in which he is described as tall. The value for n, against another larger number, represents a lower reasonable credence in a state of affairs than the higher number. Conditional probability distributions will be formed over a range of heights.

Types and Situation Types A more recent approach to situation theoretic semantics is systematically laid out in [15] which adopts a rich proof theoretic type theory. Rather than treating types as sets of objects in a domain, this approach treats types as objects (in a domain of types). Situations/events then act as proofs of propositions. In what is to follow, I will adopt a fairly simple type theory with an, essentially, model theoretic semantics.[14] This simplification will make it easier to focus on the probabilistic element in my account. A proof theoretic approach with a richer type theory could, in principle, adopt the core ideas in the proposal that I will make. Indeed, a probabilistic approach comparable to my own has just been proposed in [16].[15]

Properties It is important to note that it is not assumed that there is anything like the property *green* or the property *tall* that individuals have or lack. We will be assuming that individuals can have the property of being some height (this will just be the height that they are). We will also be assuming that certain objects have some hue or shade.[16] These heights and shade properties will be mentioned in the formalism as properties of objects in infons.

---

[14] The hedging here is included because, although I assume a domain of situations (as a set), situations are best not viewed as a set.

[15] A difference between this, my own, and similar proposals will be explained in §7.

[16] I assume for simplicity that these properties are not vague. Nothing turns on this however. Even if properties were vague in the sense that someone may genuinely be of indeterminate (exact) height, one would still need to explain how words like 'tall' admit of wide ranges of those properties in a graded way.

4.2 Definitions

Types The type system that will be used will have three basic types.[17] Type e will be the standard type for individuals (entities). Type s will be the type for situation. Rather than the traditional type t (for truth value), we will have type p (for probability). Basic types can be used to form functional types in the way normal to model theory.

Definition 1: Types

Basic types:	$\{e, p, s\}$
Functional types:	If a, b are types, then $\langle a, b \rangle$ is a type.

Expressions This formalism brings together notions from both probability theory and situation theory. As such, we will need to define how various parts of them interact. The vocabulary of the formalism is in Definition 1. Where necessary, types of expressions will be given as subscripts. For example, a variable, S, of type $\langle e, p \rangle$ will be written $S_{\langle e,p \rangle}$. The marking of numerical variables as type p in Definition 1 is an oversimplification. These will be interpreted as numerical normalising values. These values can have a value outside of $[0, 1]$, but are derivable from sums of values of interpretations of other type p expressions.

Definition 1: Vocabulary

Parameters (PAR):	infinite set of parameters for all types $\dot{s}_1, ..., \dot{d}_1, ..., \dot{x}, ...$
Constants (CON):	a possibly empty set of constants for every type.
Supports:	$\models$
Conditional on:	$\mid$
Infons:	σ, τ, $\langle\langle \texttt{height}{=}h, \texttt{j}, yes \rangle\rangle$, $\langle\langle utters, \texttt{a}, \text{TALL}, \texttt{j}, no \rangle\rangle$
Connectives:	$\wedge, \neg, \vee$
Lambda abstraction operator:	λ
Mathematical Operators:	$\times, +, -$
Numerical Variables:	$\mathbf{C}, \mathbf{C}'$ of type $\langle p \rangle$
Brackets:	(,), [,]

A further few remarks on infons are needed at this juncture. Infons contain relations between objects and the polarities in infons '*yes*' and '*no*' indicate whether the relation holds between the objects. I assume a basic ontology of relations (such as being some height or being some shade). Propositions state that situations support infons. Positive polarities on infons indicate that any situation that supports the infon is of that type. A negative polarity would indicate that the situation is of the negative type. Infons can thus be understood

[17] I will from now on suppress all reference to locations and times, which are also types in situation theory.

as types of situations [13], or perhaps as properties of situations. The class of infons will be formed of all possible combinations of relations and objects in our ontology (plus a polarity).

The well-formed expressions of the language are defined in Definition 2 (where $\mathrm{WE_a}$ denotes well formed expressions of type a). Clauses (i), (ii), (iv) and (v) are standard. Clause (v) introduces the syntax for situation types. Clause (vi) gives what will eventually be interpreted as a conditional probability of one situation type given another. The connectives in clause (vii) will be discussed in greater detail below. Clause (viii) allows for expressions of type p to be linked by mathematical connectives.

Definition 2: Well-Formed Expressions:

(i) If $\alpha \in \mathrm{PAR_a}$, then $\alpha \in \mathrm{WE_a}$,
(ii) If $\alpha \in \mathrm{CON_a}$, then $\alpha \in \mathrm{WE_a}$,
(iii) If $\alpha \in \mathrm{WE_{\langle a,b \rangle}}$ and $\beta \in \mathrm{WE_a}$, then $\alpha(\beta) \in \mathrm{WE_b}$,
(iv) If $\alpha \in \mathrm{WE_b}$, and $\beta \in \mathrm{WE_a}$, then $\lambda[\beta](\alpha) \in \mathrm{WE_{\langle a,b \rangle}}$,
(v) If $\alpha \in \mathrm{WE_s}$, and if σ is an infon, then $\lambda[\alpha](\alpha \vDash \sigma) \in \mathrm{WE_p}$,
(vi) If $\phi, \psi \in \mathrm{WE_p}$, then $\phi \mid \psi \in \mathrm{WE_p}$,
(vii) If $\phi, \psi \in \mathrm{WE_p}$, then:
  (a) $\neg\phi \in \mathrm{WE_p}$,
  (b) $\phi \wedge \psi \in \mathrm{WE_p}$,
  (c) $\phi \vee \psi \in \mathrm{WE_p}$,
(viii) If $\phi, \psi \in WE_p$, then
  (a) $\phi \times \psi \in \mathrm{WE_p}$,
  (b) $\phi + \psi \in \mathrm{WE_p}$,
  (c) $\phi - \psi \in \mathrm{WE_p}$,

As mentioned briefly earlier, there are two kinds of situation parameters that we will use: *described situations* ($\dot{s}$) and *discourse situations* ($\dot{d}$). For descriptive situation types, we will be interested in how the world is with respect to things like the shades things are (for colour terms), and the heights things are (for terms like 'tall'). In discourse situation types, what will be described is the production of some utterance type.

The infons we will be concerned with are likewise of two kinds. Infons for descriptive situations will describe/be about, say, what height someone is. Immediately below I specify the infon, and below that, I give a described situation type including that infon:

$$\langle\!\langle \texttt{height=180cm}, \texttt{j}, yes \rangle\!\rangle$$

$$\lambda[\dot{s}](\dot{s} \vDash \langle\!\langle \texttt{height=180cm}, \texttt{j}, yes \rangle\!\rangle)$$

The latter is the type of situation in which the individual denoted by `j` (say, John) is 180 cm in height. Infons for descriptive situation types will typically relate to what has been said. For example, below, I specify a discourse infon, followed by a discourse situation type including that infon:

$$\langle\langle \texttt{utters}, \dot{a}, \text{TALL}, \texttt{j}, yes \rangle\rangle$$

$$\lambda[\dot{d}](\dot{d} \vDash \langle\langle \texttt{utters}, \dot{a}, \text{TALL}, \texttt{j}, yes \rangle\rangle)$$

The latter is the type of situation in which the individual denoted by j (say, John) has been described as tall. This is a simplification, however. In actual fact, there ought to be unbound parameters for locations and times etc. which are being suppressed for simplicity.

For either of these situation types (descriptive and discourse), many actual situations could be of those types (types where John is 180 cm in height, or where he is described using 'tall').

Interpretations and Domains The interpretation function we shall use will be marked as: $p(\cdot)$. Hence, an expression ϕ of type p, will be interpreted as $p(\phi)$, or the probability of ϕ. Like interpretation functions in model theory, the interpretations of type p expressions will be built up of the interpretations of other expressions. I assume a domain for each basic type. If a is a basic type:

$$p(\alpha_{\langle a \rangle}) \in Dom_{\langle a \rangle}$$

Functionally typed expressions will be interpreted as functions from basic domains:

$$p(\alpha_{\langle a,b \rangle}) \in Dom_{\langle b \rangle}^{Dom_{\langle a \rangle}}$$

The type e domain will be a set of individuals. The domain of type p will be the range $[0, 1]$. Situations are being taken as primitive. We can, for the sake of the formalism, assume a domain of situations. However, in the spirit of situation theory, situations should be taken to be our ways of conceptualising and carving up the world into parts. As well as the interpretation function, we will have a parameter anchoring function g, which assigns appropriate members of the domain to parameters. In situation theory, parameter anchoring functions have restrictions on their domains, I will pass over those details here.

The first four clauses of Definition 3 are fairly standard. Clause (v) simply states that maths operators in the object language are treated as standard when the interpretations of their operands are in the range $[0, 1]$. Clause (vi) simply states the standard definitions of probabilistic connectives [17].

Spaces and Priors In order for these axioms and the probability function to be well defined, we must define a probability space. Probabilities will distribute over situation types. The probability value will be in the range $[0, 1]$ where this value indicates the probability of some situation being of the type specified by the infon. Conditional probabilities will be the probability of a situation being of some type, given that some situation is of another type. For simplicity, for

Definition 3: Denotations

(i) $p(\dot{x})^g \;=\; g(x)$ if $\dot{x} \in PAR$,
(ii) $p(\mathtt{c})^g \;= p(\mathtt{c})$ if $\mathtt{c} \in CON$,
(iii) $p(\alpha(\beta))^g \;=\; p(\alpha)^g(p(\beta)^g)$,
(iv) $p(\lambda[\dot{x}](\alpha))^g \;=\; f$ such that if $\dot{x} \in PAR_a$, and $c \in Dom_a$, $f(c) = p(\alpha)^{g[\dot{x}:=c]}$,
(v) Maths Operators:
  (a) $p(\phi \times \psi)^g \;=\; p(\phi)^g \times p(\psi)^g$,
  (b) $p(\phi + \psi)^g \;=\; p(\phi)^g + p(\psi)^g$,
  (c) $p(\phi - \psi)^g \;=\; p(\phi)^g - p(\psi)^g$,
(vi) Probabilistic Connectives:
  (a) $p(\neg\phi)^g \;=\; 1 - p(\phi)^g$,
  (b) $p(\phi \wedge \psi)^g \;=\; p(\phi)^g \times p(\psi \mid \phi)^g$,
  (c) $p(\phi \vee \psi)^g \;=\; p(\phi)^g + p(\psi)^g - p(\phi \wedge \psi)^g$.

all non-conditional values, I assume that probabilities distribute evenly.[18] For example, in the case where we have n discrete ranges of heights over which a distribution will be formed, the values for each unconditional situation type will be:

$$p(\lambda[\dot{s}](\dot{s} \vDash \langle\!\langle \mathtt{height}\!=\!\mathtt{h}, \dot{x}, yes\rangle\!\rangle)^g = \frac{1}{n}$$

In practice, the range of situations and infons will be highly constrained. For described situations, constraints will come out of the semantic learning process, as well as the goals and purposes of the speakers. For example, certain ranges of shade/hue properties will constrain the range of described situations for colour terms in general, but further constraints may come from what objects are salient to speaker and hearer in the situation of utterance.

For discourse situations, I assume a space of two infons, formed of the same relation and object but with opposing polarities.

$$p(\lambda[\dot{d}](\dot{d} \vDash \langle\!\langle \mathtt{R}, \mathtt{a_1}, ...\mathtt{a_n}, yes\rangle\!\rangle))^g \;=\; 0.5 \;=\; p(\lambda[\dot{d}](\dot{d} \vDash \langle\!\langle \mathtt{R}, \mathtt{a_1}, ...\mathtt{a_n}, no\rangle\!\rangle))^g$$

However, when multiple terms are used, the space of possibilities may be bigger and so values may be lower.

Representationalism The formal semantics and linguistics debate over representationalism got going with the publication of work on Discourse Representation Theory (DRT) [18]. DRT questioned Montague's claim that the representational language of a formalism is dispensable. In DRT, it was claimed that to account for the treatment of some expressions (such as discourse anaphora), the

[18] In a more sophisticated model, one would wish the values of priors to be set in accordance with the learning experiences of agents. One way this has been implemented will be discussed in §7.2.

representational (interpretational) level is not dispensable. Non-representational, dynamic accounts were developed in response. The debate was never entirely settled, but I do not need to take up a position.

Linguistic Meaning In standard situation semantics, linguistic meaning is captured by way of *constraints*. For example, the linguistic meaning of 'Mary is tall' will be a link between an utterance situation type in which an utterance of 'Mary is tall' is made, and a described situation type in which Mary is tall. The introduction of probabilities into the situation theoretic framework, can be understood as the loosening of the strength of this link. Rather than taking an utterance of 'Mary is tall' to convey information in such a way as to give us the expectation that in some described situation Mary is tall, probabilistic constraints will hold between utterance situation types and a number of described situation types. In each of the described situation types, Mary will be some particular height, however, the strength with which we should expect Mary to be that height, given that she has been described as tall will vary. This uncertainty will be captured formally as a conditional probability distribution. For the example in hand this will be the following:

$$p(\lambda[\dot{s}](\dot{s} \vDash \langle\langle \texttt{height}{=}h, \texttt{m}, yes \rangle\rangle) \mid \lambda[\dot{d}](\dot{d} \vDash \langle\langle \texttt{Utters}, \dot{a}, \text{TALL}, \texttt{m}, yes \rangle\rangle))^{g}$$

The above formula will return probability values for different values of heights h. Put simply, given a type of situation in which Mary is described as tall, one should assign a greater probability to the described situation being one in which Mary is some heights rather than others. However, these values will be determined by more general correlations in which, say, people are described as tall and where people are some height or other (the details of this will emerge in the compositional semantics). The standard way we will represent and interpret declarative sentences will therefore be as a conditional probability formed with a described situation type and a discourse situation type where the resultant probability distribution describes the constraints that the words used in the utterance situation place on the described situation:

$$p(\lambda[\dot{s}](\dot{s} \vDash \tau) \mid \lambda[\dot{d}](\dot{d} \vDash \sigma))^{g}$$

For different infons substituted for τ, values will be in the range $[0, 1]$. These values will form a distribution. Returning to our example of 'Mary is tall', using some made-up values, the sentence could be interpreted as in Table 1. In §5, I will describe how such a sentence can be composed.

5 Semantics: Terms

5.1 Predicates

Since nominal predicates were argued to be important for the semantics of vague adjectives/modifiers (as the things bearing the information that they modify),

Table 1. Interpretation of 'Mary is tall'

h (cm)	$p(\lambda[\dot{s}](\dot{s} \vDash \langle\!\langle \texttt{height}{=}h, \texttt{m}, yes\rangle\!\rangle) \mid \lambda[\dot{d}](\dot{d} \vDash \langle\!\langle \texttt{Utters}, \dot{a}, \text{TALL}, \texttt{m}, yes\rangle\!\rangle))$
$h < 150$	0.01
$150 \leq h < 155$	0.01
$155 \leq h < 160$	0.01
$160 \leq h < 165$	0.01
$165 \leq h < 170$	0.01
$170 \leq h < 175$	0.02
$175 \leq h < 180$	0.06
$180 \leq h < 185$	0.11
$185 \leq h < 190$	0.22
$190 \leq h < 195$	0.32
$h > 195$	0.22

we will now turn to them. As suggested in §3, nominal predicates may carry a lot of different information. Our basic semantic representation for such predicates will incorporate an argument the domain of which will be a selection of a type of information (such as a range of heights). This information type will be supplied either by context, or by a modifier, such as an adjective. In a departure from a simple view of predicates (that take just an object as an argument), nominal predicates will be modelled as a function from properties to a function from an individual to a probability, which is to say that they will be a function from properties to properties $(\langle e, p\rangle, \langle e, p\rangle)$. The logical structure of nominal expressions will be such that two entities in the above will be the same. Put another way, singular nominal predicates are assumed to require updating in context with respect to the aspect of the information they carry (via an argument of type $\langle e, p\rangle$). Then, when provided with an individual (a type e argument), yield a value for the probability of that individual having that property, given the information provided. The schema for 'man' (and with appropriate substitution, other predicates) is:

$$\text{Man}: \ \lambda[\dot{S}](\lambda[\dot{x}](\dot{S} \ . \ [\dot{x}] \mid \lambda[\dot{d}](\dot{d} \vDash \langle\!\langle \texttt{Utters}, \dot{a}, \text{MAN}, \dot{x}, yes\rangle\!\rangle)))_{\langle e,p\rangle,\langle e,p\rangle}$$

$\dot{x}, \dot{y}$:= Parameters for individuals $(\langle e\rangle)$
$\dot{S}, \dot{R}$:= Parameters of functional type $\langle e, p\rangle$
e.g. $\lambda[\dot{y}](\lambda[\dot{s}](\dot{s} \vDash \langle\!\langle \texttt{height}{=}\texttt{180cm}, \dot{y}, yes\rangle\!\rangle))$

To get the information carried by 'is a man' with respect to the having of some property, we need to provide a property a man might have. Such a property (or range of properties) may be provided by the context. Below, however, we will see how part of the semantic contribution of predicate modifiers is to provide such a property (such a range of properties).

Because nominal predicates are functions from properties of described situation types to a function from individuals to a probability, a worry over tractability and learnability arises. If an agent needs to learn all distributions that arise from interpreting 'x is a man' constructions, this would not be a tractable learning task. However, probabilistic learning affords a solution. We can assume that agents begin classifier learning tasks with flat prior distributions for all collections of properties. On being exposed to uses of a classifier, distributions will be adjusted for just those properties the objects are perceived to have. In this sense, if one has no reasonable expectations regarding some range of situations, given what has been said, simply because no one situation type is more plausible than another, then it is possible that the classifier used does not carry information about (the properties in) those situations.[19]

One property a man might have is of being a certain height. For example, he might be 180cm in height:

$$\lambda[\dot{y}](\lambda[\dot{s}](\dot{s} \vDash \langle\langle \texttt{height=180cm}, \dot{y}, yes\rangle\rangle))$$

This kind of property can be applied to our representation of the predicate 'man':

$$\begin{aligned}
&\lambda[\dot{S}](\lambda[\dot{x}](\dot{S}\ .\ [\dot{x}] \mid \lambda[\dot{d}](\dot{d} \vDash \langle\langle \texttt{utters}, \dot{a}, \text{MAN}, \dot{x}, yes\rangle\rangle)\ .\\
&\quad [\lambda[\dot{y}](\lambda[\dot{s}](\dot{s} \vDash \langle\langle \texttt{height=180cm}, \dot{y}, yes\rangle\rangle)]\\
\Rightarrow&\lambda[\dot{x}](\lambda[\dot{y}](\lambda[\dot{s}](\dot{s} \vDash \langle\langle \texttt{height=180cm}, \dot{y}, yes\rangle\rangle)\ .\ [\dot{x}] \mid\\
&\quad \lambda[\dot{d}](\dot{d} \vDash \langle\langle \texttt{utters}, \dot{a}, \text{MAN}, \dot{x}, yes\rangle\rangle)))\\
\Rightarrow&\lambda[\dot{x}](\lambda[\dot{s}](\dot{s} \vDash \langle\langle \texttt{height=180cm}, \dot{x}, yes\rangle\rangle \mid \lambda[\dot{d}](\dot{d} \vDash \langle\langle \texttt{utters}, \dot{a}, \text{MAN}, \dot{x}, yes\rangle\rangle)
\end{aligned}$$

This is now in the right shape to take an individual as an argument which will be described as a man in $\dot{d}$ and assigned the height property in $\dot{s}$. Assuming (as a simplification) that names refer and carry no other information, this means that 'John is a man' can be represented, with respect to being 180cm in height, as follows:[20]

$$\lambda[\dot{s}](\dot{s} \vDash \langle\langle \texttt{height=180cm}, \texttt{j}, yes\rangle\rangle) \mid \lambda[\dot{d}](\dot{d} \vDash \langle\langle \texttt{utters}, \dot{a}, \text{MAN}, \texttt{j}, yes\rangle\rangle)$$

The interpretation of this will be a probability value that reflects, with respect to John having some height, the probability of John having that height, given that he has been described as a man.

Of course, 'John is a man' will carry more information than this about John, but let us focus on information carried about his height. We can consider the

[19] Of course, given particularly skewed learning data, anomalies in individuals semantic representations may occur. There will, nonetheless, be general patterns of use across whole language communities. What learners are assumed to be implicitly doing is approximating to the patterns of use in their learning communities as a whole.

[20] I ignore the contribution of the indefinite article and treat this as a simple predication. I leave the modelling of quantification in this system for future research.

range of heights John might be (given that he has been described as a man). There will be a rationality constraint on the values 'John is a man' receives with respect to John's height. The values, must sum to 1. As a simplification, we may think in discrete values, although a more accurate representation would be a continuous function.

Put another way, providing different arguments of height properties for the formula above will generate a probability distribution. Table 2 shows one possible distribution for the height information carried by 'John is a man' simplified over ranges of heights.

Table 2. Interpretation of 'John is a man' with respect to height.

h (cm)	$p(\lambda[\dot{s}](\dot{s} \vDash \langle\langle \texttt{height}{=}h, \texttt{j}, yes \rangle\rangle) \mid \lambda[\dot{d}](\dot{d} \vDash \langle\langle \texttt{utters}, \dot{a}, \text{MAN}, \texttt{j}, yes \rangle\rangle))$
$h < 150$	0.01
$150 \leq h < 155$	0.04
$155 \leq h < 160$	0.08
$160 \leq h < 165$	0.12
$165 \leq h < 170$	0.15
$170 \leq h < 175$	0.20
$175 \leq h < 180$	0.15
$180 \leq h < 185$	0.12
$185 \leq h < 190$	0.08
$190 \leq h < 195$	0.04
$h > 195$	0.01

Describing John as a man, in part, conveys information about his height: that it is fairly probable that he is around average height and highly probable that he is neither far below nor far above average height. This relates to the assumption that learning the classifier 'man', in part, involves learning, approximately, the ranges of heights men tend to come in. This needn't be in centimetres. It need only be internalised in such a way as to aid, to some extent, the ability to identify and classify whether objects in one's environment are men.

Tall As a predicate modifier, 'tall', when applied to a nominal predicate, increases our (reasonable) expectations of the entity (that predicate is applied to) having a greater height. 'Tall' will therefore be modelled as having two jobs to do when applied to a common noun like 'man': (i) it will pick out the information carried by 'man' with respect to height. (ii) it will be a function on the probability distribution that 'man' generates with respect to heights (it will make taller heights more probable and shorter heights less probable). In terms of a distribution curve, for a graph with probabilities on its y-axis and heights on its x-axis, it would shift the whole curve along the x-axis in the direction of greater heights. The representation of 'tall' is given below, along with a derivation for 'tall man'.

$$
\begin{array}{ll}
\textit{tall}: & \lambda[\dot{P}](\lambda[\dot{y}](\mathbf{C} \times f_{\text{tall}}((P\ .\ [\dot{y}])\ .\ [(\lambda[\dot{z}](\dot{s} \vDash \langle\!\langle \texttt{height=h}, \dot{z}, yes \rangle\!\rangle)]))))\\
\textit{man}: & \lambda[\dot{S}](\lambda[\dot{x}]((\dot{S}\ .\ [\dot{x}] \mid \dot{d} \vDash \langle\!\langle \texttt{utters}, \dot{a}, \text{MAN}, \dot{x}, yes \rangle\!\rangle)))\\
\textit{tall man:} & \lambda[\dot{P}](\lambda[\dot{y}](\mathbf{C} \times f_{\text{tall}}((\dot{P}\ .\ [\dot{y}]))\ .\ [(\lambda[\dot{z}]\dot{s} \vDash \langle\!\langle \texttt{height=h}, \dot{z}, yes \rangle\!\rangle)])))\ .\\
& \quad [\lambda[\dot{S}]\lambda[\dot{x}]\ (\dot{S}\ .\ [\dot{x}] \mid \dot{d} \vDash \langle\!\langle \texttt{utters}, \dot{a}, \text{MAN}, \dot{x}, yes \rangle\!\rangle)]\\
& \Rightarrow \lambda[\dot{y}](\mathbf{C} \times\ f_{\text{tall}}(\ \lambda[\dot{s}](\dot{s} \vDash \langle\!\langle \texttt{height=h}, \dot{y}, yes \rangle\!\rangle) \mid\\
& \quad \lambda[\dot{d}](\dot{d} \vDash \langle\!\langle \texttt{utters}, \dot{a}, \text{MAN}, \dot{y}, yes \rangle\!\rangle))))
\end{array}
$$

$\dot{P}$:= Parameter of type $\langle\langle e,p\rangle,\langle e,p\rangle\rangle$ (functional type from situation properties to a function from individuals to $[0,1]$).
C := Normalising value.
f_{tall} := Expression of type $\langle p,p\rangle$.

The interpretation of which is a function from individuals to a probability distribution for heights of that individual, given their description as a tall man. Whatever function on distributions f_{tall} is, **C** will be interpreted as the value that normalises the distribution.[21] One possible interpretation for f_{tall} could be that it has the effect of shifting the distribution for 'man' up by 20cm or so (and stretching it vertically).[22] In Table 3, from the above toy distribution for 'John is a man', we can get a toy distribution for 'John is a tall man'.[23]

Table 3. Interpretation of 'man', and 'tall man' (unnormalised and normalised) with respect to height

h (cm)	$p(\Phi)$	$p(f_{\text{tall}}(\Phi))$	$p(\mathbf{C} \times f_{\text{tall}}(\Phi))$
$h < 150$	0.01	0.01	0.01
$150 \leq h < 155$	0.04	0.01	0.01
$155 \leq h < 160$	0.08	0.01	0.01
$160 \leq h < 165$	0.12	0.01	0.01
$165 \leq h < 170$	0.15	0.01	0.01
$170 \leq h < 175$	0.20	0.02	0.02
$175 \leq h < 180$	0.15	0.06	0.06
$180 \leq h < 185$	0.12	0.10	0.11
$185 \leq h < 190$	0.08	0.20	0.22
$190 \leq h < 195$	0.04	0.30	0.32
$h > 195$	0.01	0.20	0.22

[21] Which will be 1 over the sum of the modified distribution.
[22] This could also be described by adjusting values of parameters on a Gaussian function.
[23] I assume that 0.01 is the arbitrarily small value. Φ is an abbreviation for $\lambda[\dot{s}](\dot{s} \vDash \langle\!\langle \texttt{height=}h, \texttt{j}, yes \rangle\!\rangle \mid \dot{d} \vDash \langle\!\langle \texttt{utters}, \dot{a}, \text{MAN}, \texttt{j}, yes \rangle\!\rangle)$

Simply translated, being told that John is a man may carry all sorts of information. Some of this is that he is more likely than not to be within some margin of average height. Being told that he is tall carries the information that it is highly probable that John is over average height. 'tall' selects a particular kind of information carried by its common noun (namely information about height), and it amplifies expectations for the upper bounds.

Importantly, after normalisation, we are simply left with a probability distribution (in the final column). Any of those values will be able to enter into formulas for further manipulation if need be. It will be convenient to mark, in the syntax of the formalism, that normalised distributions are mere distributions. This can amount to a dropping of the f and $\mathbf{C}$, but keeping a record of what the normalised distribution is over. Hence, the interpretation of formulas like:

$$\mathbf{C} \times \; f_{\text{tall}}(\; \lambda[\dot{s}](\dot{s} \vDash \langle\!\langle \texttt{height=h}, \texttt{j}, yes \rangle\!\rangle) \mid \lambda[\dot{d}](\dot{d} \vDash \langle\!\langle \texttt{utters}, \dot{a}, \text{MAN}, \texttt{j}, yes \rangle\!\rangle))$$

will, interpreted, be a probability distribution over heights, which can be rewritten with an updated discourse situation:

$$\lambda[\dot{s}](\dot{s} \vDash \langle\!\langle \texttt{height=h}, \texttt{j}, yes \rangle\!\rangle) \mid \lambda[\dot{d}](\dot{d} \vDash \langle\!\langle \texttt{utters}, \dot{a}, \text{TALL MAN}, \texttt{j}, yes \rangle\!\rangle)$$

However, in doing so we must be aware that we have increased the probability space of discourse situations. Now there are four possible discourse situation types since John might or might not be described as a man and might or might not be described as tall.

Green As noted in §3, modifiers like 'green' differ from 'tall': simply knowing that something (anything) is green *will* give reasonable expectations about what it is like. The way this will be modelled is that 'green' will also be a function on probability distributions, but whereas 'tall' was informally characterised as taking a distribution over heights and shifting it up (so that greater heights receive higher values and lower heights, lower values), 'green' will contribute something more stable to a distribution. It will select shades or hues as properties to distribute over, but it will have the effect of always weighting the distribution towards certain shades. For each shade that an object might be, 'green' provides a weighting over our expectations for what is being referred to as being one of those shades.

Formally, 'green' will look the same as 'tall' and the only difference will be in the functions f_{tall} and f_{green}. Immediately below is the derivation for 'green car'. For some shade c, the interpretation of the above is a function from individuals to a probability value. For different values of c, this will form a distribution.

Unlike 'tall', which pulls height distributions up by some factor, 'green' will have a more constant character. Both functions will be of type $\langle p, p \rangle$, but whereas f_{tall} will has the effect of moving a distribution over heights upwards (increasing expectation of heights), f_{green} will always have the effect of flattening a distribution over shades where those shades are not usually described as green, and of

green:	$\lambda[\dot{P}](\lambda[\dot{y}](\mathbf{C} \times f_{\text{green}}((\dot{P}\ .\ [y])\ .\ [\lambda[\dot{z}](\lambda[\dot{s}](\dot{s} \vDash \langle\!\langle \texttt{shade=c}, \dot{z}, yes \rangle\!\rangle)]]))))$
car:	$\lambda[\dot{S}](\lambda[\dot{x}](\dot{S}\ .\ [\dot{x}] \mid \lambda[\dot{d}](\dot{d} \vDash \langle\!\langle, \texttt{utters}, \dot{a}, \text{CAR}, \dot{x}, yes \rangle\!\rangle)))$
green car:	$\lambda y.\mathbf{C} \times f_{\text{green}}(\lambda[\dot{s}](\dot{s} \vDash \langle\!\langle \texttt{shade=c}, \dot{y}, yes \rangle\!\rangle) \mid \lambda[\dot{d}](\dot{d} \vDash \langle\!\langle \texttt{utters}, \dot{a}, \text{CAR}, \dot{y}, yes \rangle\!\rangle))$

elevating the distribution over shades that are usually described as green. The function for 'green' will pull any predicate distribution it is applied to towards the same points, namely, some range of shades. This will represent why it is more reasonable to expect something described as 'green' to be some shades. This also accounts for why 'green' seems to convey information on its own (about shades) in a way that 'tall' does not (about heights).

5.2 Connectives

For the interpretations of two declarative statements:

$$p(\lambda[\dot{s}](\dot{s} \vDash \sigma) \mid \lambda[\dot{d}](\dot{d} \vDash \tau))$$
$$p(\lambda[\dot{s}](\dot{s} \vDash \sigma) \mid \lambda[\dot{d}](\dot{d} \vDash \rho))$$

The interpretation of their conjunction and disjunction will be:

$$p(\lambda[\dot{s}](\dot{s} \vDash \sigma) \mid \lambda[\dot{d}](\dot{d} \vDash \tau) \wedge \lambda[\dot{d}](\dot{d} \vDash \rho))$$
$$p(\lambda[\dot{s}](\dot{s} \vDash \sigma) \mid \lambda[\dot{d}](\dot{d} \vDash \tau) \vee \lambda[\dot{d}](\dot{d} \vDash \rho))$$

Values for which can be obtained using Bayes' Rule.[24] For example, in the conjunction case, this would yield:

$$p(\lambda[\dot{s}](\dot{s} \vDash \sigma) \mid \lambda[\dot{d}](\dot{d} \vDash \tau) \wedge \lambda[\dot{d}](\dot{d} \vDash \rho)) = \frac{p(\lambda[\dot{s}](\dot{s} \vDash \sigma)) \ \wedge \ p(\lambda[\dot{d}](\dot{d} \vDash \tau) \wedge \lambda[\dot{d}](\dot{d} \vDash \rho))}{p(\lambda[\dot{d}](\dot{d} \vDash \tau) \wedge \lambda[\dot{d}](\dot{d} \vDash \rho))}$$

[24] Bayes' Rule, stated in regular notation is: $P(C|A) = \frac{P(C \wedge A)}{P(A)}$.

6 Vagueness

On the semantics being developed, a phenomena (one that might well be called 'vagueness'), emerges naturally from meaning representations. If terms leave us with U2 uncertainty, then sometimes there will be a clash between our intuitions of whether to apply a particular term in a particular case. U2 uncertainty (uncertainty over how to apply terms) is related to U1 uncertainty (uncertainty over what the world is like given a description). If U1 uncertainty is graded, then there will undoubtedly arise cases where U2 uncertainties clash: cases where we are no more certain about applying a term than we are about applying its negation. Furthermore, some cases of clashes will be immune to further information. One could be in a position of full knowledge such that one knew exactly how words were used, all of the relevant properties of the objects being described, and all of the relevant contextual information. Nonetheless, clashes of equal uncertainty may remain unresolved. Such situations, I contend, are plausible explanations of so-called borderline cases.

For example, if the value for 'John is tall' with respect to height h is the same as the value for 'John is not tall' with respect to height h, then the meaning of 'tall' will provide us with no more reason to judge John to be tall than to judge him to be not tall. There is no more to know about the meaning of 'tall' which will resolve this uncertainty, and there may be no more to know about John or the context of utterance to do so either. If so, John would be a borderline case.

Probability values in the formalism can be seen to reflect reasons for making judgements. In viewing matters this way, we can begin to get a handle on boundarylessness. Even if our reasons for making a 'tall' judgement slightly outweigh our reasons for making a 'not tall' judgement, we might still not have *sufficient* reason to form a categorical judgement. The extent to which the values must differ for one or the other judgement to be true will not be a matter decided by the semantics for 'tall'. In this sense, nothing in the meaning representation itself will determine a cut-off point for 'tall'. It is due to this feature of the model provided that we can approximate boundarylessness for vague concepts.

7 Comparison with the Literature

7.1 Lassiter's, and Frazee and Beaver's Epistemic Models

Although [2] and [3] describe their positions from different perspectives, they amount to similar approaches. Lassiter's article, though marginally later was developed independently and is fuller in detail. I will focus on it.

Similarly to my approach, Lassiter is interested in uncertainty in communication and worried about sharp boundaries. He is less worried about truth-conditional semantics in general, however. Rather than implementing probabilities directly into the meaning representations of words, Lassiter incorporates metalinguistic uncertainty over what is being expressed (in terms of a sharp proposition). He also includes uncertainty about the world, but we'll focus on the former case.

Lassiter inherits, from other work in the scalar adjectives literature, the idea that their semantics should incorporate thresholds. In simple terms, he models uncertainty about where the threshold is when a term like 'tall' is used. We can, relative to a context, be pretty certain that thresholds are not too high or too low. Lassiter models this as a distribution over precise model theoretic objects which have sharp cut-offs. The effect is that if we learn someone's height, we can get a value that reflects the probability that that individual is tall.

However, Lassiter's emphasis is not on the sharp boundaries that such a position implies. For him (and for Frazee and Beaver), communication using 'tall' is about approximating, roughly, the standard in play for what would count as tall. It is these standards that we are uncertain about when someone uses the expression 'tall'. However, this general picture has the effect of implying that there is, nonetheless, a standard for 'tall' in every context. I have suggested that the approximation of standards is not what is encoded by our words. This amounts to the claim that uncertainty should enter at the level of describing the model theoretic object, not at the level of evaluating which classical model theoretic object is in play.

Lassiter's account is a big improvement on standard non-probabilistic approaches. However, the fact remains that the motivation for appealing to precise languages must still be given. If it turns out that we rarely, if ever, coordinate sufficiently to settle on exactly where a threshold for a term for 'tall' should be, then why have such thresholds written so centrally into the semantics of such terms? We can drop precise languages and have a far more direct connection to what our terms mean by adopting the picture proposed.

7.2 Cooper et al.'s Probabilistic Type-Theoretic Approach

Semantic learning is at the forefront of [16], which is also the closest position to my own. I strongly suspect that the core of the two positions will be pretty inter-definable, although there are differences of emphasis.

U1 uncertainty is uncertainty about the world, given a description of it. U1 uncertainty is, effectively, what my account describes. Arguably, Cooper *et al.* focus on U2 uncertainty. U2 uncertainty is uncertainty about how to use words, given some known or perceived way the world is. I will say more about Cooper *et al.*'s formalism in a moment, but it is first worth remarking that, if modelled as I have presented, U1 and U2 uncertainty are inter-definable. The main weight of the meanings of utterances in my account rests on conditional probabilities of the form $p(\lambda[\dot{s}](\dot{s} \vDash \sigma)|\lambda[\dot{d}](\dot{d} \vDash \tau))$. In other words, the probability of the world being some way, given a description of it. If, indeed, Cooper *et al.*'s account describes U2 uncertainty, then it is possible that their results could be simulated via the use of the alternate conditional probability: $p(\lambda[\dot{d}](\dot{d} \vDash \tau)|\lambda[\dot{s}](\dot{s} \vDash \sigma))$ (or the probability that some description will be used, given that the world is some way). Importantly, given the priors $p(\lambda[\dot{s}](\dot{s} \vDash \sigma))$ and $p(\lambda[\dot{d}](\dot{d} \vDash \tau))$, these two conditional probabilities are simply ratios of each other.

Cooper *et al.*'s semantics uses a rich theory of types. The simple type theory I have used provided domains for basic types. Complex types were then constructed as functions of basic types. In rich type theories, types should not be thought of in these extensional terms, but instead as something like useful ways of classifying things [15, p. 275]. However, propositions are also types. For example, the proposition *David runs* can be seen as a situation type (one in which David runs), and is true iff there is a situation in the world in which David runs. In [16] agents are modelled as making probabilistic judgements about classifications. Propositions are still types, but judgements reflect the probabilities that there is a situation of the right type.

Of particular interest is how these type judgements are grounded. Central to Cooper *et al.*'s account is semantic learning, a wider discussion of which features in [19]. On their learning model, a learner is exposed to multiple situations. In each one they make a judgement about whether the object they are shown is an apple or not an apple. After each judgement, the oracle that models the adult speaker gives a judgement. Initially, the learner cannot make a judgement, but after a few exposures to adult judgements, has enough data to begin to make judgements themselves. In Cooper *et al.*'s model, 'apple' judgements are based on four properties (two colours and two shapes). Simplifying a little, when faced with a new object to classify, the value of the judgement is calculated as the conditional probability $p(apple|observed\ properties)$.[25] To calculate that conditional probability, the learner uses priors and conditional probabilities such as $p(property|apple)$, both of which are estimated directly from the adult judgements they have witnessed.

Translating more into my own vernacular, one could see probabilities of 'apple' judgements as approximating probabilities of discourse situations, and we could see probabilities of properties as approximating probabilities of described situations. The two stages of Cooper *et al.*'s learning account can then be described with the following procedure:

(i) By directly witnessing adult speakers' linguistic behaviour, estimate probabilities of the form:
$p(\lambda[\dot{s}](\dot{s} \vDash \langle\langle colour_c, \dot{x}, yes\rangle\rangle))$, $p(\lambda[\dot{s}](\dot{s} \vDash \langle\langle shape_s, \dot{x}, yes\rangle\rangle))$,
$p(\lambda[\dot{d}](\dot{d} \vDash \langle\langle \texttt{utters}, \dot{a}, \text{APPLE}, \dot{x}, yes\rangle\rangle))$,
$p(\lambda[\dot{s}](\dot{s} \vDash \langle\langle colour_c, \dot{x}, yes\rangle\rangle)|\lambda[\dot{d}](\dot{d} \vDash \langle\langle \texttt{utters}, \dot{a}, \text{APPLE}, \dot{x}, yes\rangle\rangle))$, and
$p(\lambda[\dot{s}](\dot{s} \vDash \langle\langle shape_s, \dot{x}, yes\rangle\rangle)|\lambda[\dot{d}](\dot{d} \vDash \langle\langle \texttt{utters}, \dot{a}, \text{APPLE}, \dot{x}, yes\rangle\rangle))$.

(ii) Use those values to calculate, of a novel object/context, the probability:
$p[\lambda[\dot{d}](\dot{d} \vDash \langle\langle \texttt{utters}, \dot{a}, \text{APPLE}, \dot{x}, yes\rangle\rangle))|$
$\lambda[\dot{s}]\dot{s} \vDash \langle\langle colour_c, \dot{x}, yes\rangle\rangle \wedge \lambda[\dot{s}]\dot{s} \vDash \langle\langle shape_s, \dot{x}, yes\rangle\rangle)$.

The process in (i) describes learning the kinds of meaning representations I have proposed. Cooper *et al.* show how the kind of information that can be learnt from adult speakers' linguistic behaviour can be turned into a classifier judgement. These judgements will not always be probability 1. The uncertainty that < 1 judgements reflect, is, I would argue, just what I described as U2

[25] Where the output judgement is decided by whether an 'apple' or an 'not-apple' judgement receives a higher value.

uncertainty. A closer examination of the two accounts could be fruitful. It would certainly be useful if I were simply able to adopt Cooper *et al.*'s learning model.

One distinct contribution the account in this paper makes is in the treatment of adjectives. Cooper *et al.*'s learning account shows how a nominal like 'apple' could be learnt from basic colour and shape observations. I have argued that vague modifiers should be viewed as functions on distributions given by nominal classifiers.

8 Conclusions

Situation theory captures sentence meaning as a link between discourse situation types and described situation types. Although the way that this idea has been developed here differs from the situation theoretic account in many ways, the basic spirit of the account remains. On the standard view, informational links hold between situation types. Individuals build connections with the world via learning (learning to decode the information carried by terms) and via communication. The reproduction of expressions for similar purposes entrenches these relations.

On the account presented here, the meanings of expressions are correlational informational links between discourse situation types and described situation types that arise from patterns in language use. People connect to these correlations via semantic learning, and they entrench correlations via reproduction of language to refer to similar situations.

Vagueness naturally arises from modelling such relations probabilistically. This can be viewed either as the boundarylessness of vague predicates achieved by separating, or at least distancing, truth conditions from semantic representations, or as borderline cases, where the meaning of a term such as 'tall' can provide competing reasons to make 'tall' and 'not tall' judgements.

However, the suggestions put forward here are rather programmatic and many issues have remained entirely unaddressed. Nonetheless, at least for vague terms, structured, probabilistic representations of meaning provide at least one viable route for attempting to capture the boundarylessness that seems characteristic of vagueness.

References

1. Sainsbury, R.M.: Concepts Without Boundaries, King's College London Inaugural Lecture. King's College London, London (1991). Reprinted in: Keefe, R., Smith, P. (eds.) Vagueness: A Reader, pp. 251–264. MIT Press, Cambridge, MA (1996)
2. Lassiter, D.: Vagueness as Probabilistic Linguistic Knowledge. In: Nouwen, R., Sauerland, U., Schmitz, H.C., van Rooij, R. (eds.) Vagueness in Communication. LNAI, vol. 6517, pp. 127–150. Springer, Heidelberg (2011)
3. Frazee, J., Beaver, D.: Vagueness is Rational Under Uncertainty. In: Aloni, M., Bastiaanse, H, de Jager, T., Schulz, K. (eds.) Logic, Language and Meaning: 17th Amsterdam Colloquium, Amsterdam, The Netherlands, December 16-18, 2009, Revised Selected Papers, LNAI, vol. 6042, pp. 153–162. Springer, Heidelberg (2010)

4. Sutton, P.R.: Vagueness, Communication and Semantic Information. King's College London, PhD Thesis (2013)
5. Cresswell, M.J.: The Semantics of Degree. In: Partee, B. (ed.) Montague Grammar, pp 261–292. Academic Press, New York (1974)
6. Gargett, A., Gregoromichelaki, E., Kempson, R., Purver, M., Sato, Y.: Grammar resources for modelling dialogue dynamically. Cognitive Neurodynamics 3, 347–363 (2009)
7. de Marnee, M., Manning, C.D., Potts, C.: "Was it good? It was provocative." Learning the meaning of scalar adjectives. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 167–176. ACL, Stroudsberg (2010)
8. Kamp, H.: Two Theories About Adjectives. In: Keenan, E. (ed.) Formal Semantics of Natural Language, pp. 123–155. Cambridge University Press, Cambridge (1975)
9. Kennedy, C.: Projecting the adjective: the syntax and semantics of gradability and comparison. University of California, Santa Cruz, PhD Thesis (1997)
10. Graff Fara, D.: Shifting Sands: An Interest Relative Theory of Vagueness. Philosophical Topics 28.1, 45–81 (2000)
11. Kennedy, C.: Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives. Linguistics and Philosophy. 30.1, 1–45 (2007)
12. Barwise, J., Perry, J,: Situations and Attitudes. CSLI, Stanford (1983)
13. Cooper, R., The Role of Situations in Generalised Quantifiers. In: Lappin, S. (ed.) The Handbook of Contemporary Semantic Theory, pp. 65–86. Blackwell, Cambridge MA (1996)
14. Devlin, K.: Situation Theory and Situation Semantics. In: Gabbay, D., Woods, J. (eds.) Handbook of the History of Logic, Vol. 7, pp. 601–664. Elsevier, Amsterdam (2006)
15. Cooper, R.: Type Theory and Semantics in Flux. In: Kempson, R., Fernando, T., Asher, N. (eds.) Philosophy of Linguistics (Handbook of the Philosophy of Science vol. 14), pp. 271–323. Elsevier, Oxford (2012)
16. Cooper, R., Dobnik, S., Lappin, S., Larsson, S.: A Probabilistic Rich Type Theory for Semantics Interpretation, Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics, Gothenburg (2014)
17. Kolmogorov, A.: Foundations of Probability. Chelsea Publishing, New York (1950)
18. Kamp, H., Reyle, U.: From Discourse to Logic. Kluwer, Dordrecht (1993)
19. Cooper, R., Dobnik, S., Lappin, S., Larsson, S.: Probabilistic Type Theory and Natural Language Semantics (manuscript) University of Gothenburg and King's College London (2013)