NICHOLAS J. J. SMITH

# VAGUENESS AND BLURRY SETS

Received 24 April 2002
Received in revised version on 25 August 2003

ABSTRACT. This paper presents a new theory of vagueness, which is designed to retain the virtues of the fuzzy theory, while avoiding the problem of higher-order vagueness. The theory presented here accommodates the idea that for any statement $S_1$ to the effect that 'Bob is bald' is $x$ true, for $x$ in [0, 1], there should be a further statement $S_2$ which tells us how true $S_1$ is, and so on – that is, it accommodates higher-order vagueness – without resorting to the claim that the metalanguage in which the semantics of vagueness is presented is itself vague, and without requiring us to abandon the idea that the logic – as opposed to the semantics – of vague discourse is classical. I model the extension of a vague predicate $P$ as a *blurry set*, this being a function which assigns a degree of membership or *degree function* to each object $o$, where a degree function in turn assigns an element of [0, 1] to each finite sequence of elements of [0, 1]. The idea is that the assignment to the sequence $\langle 0.3, 0.2\rangle$, for example, represents the degree to which it is true to say that it is 0.2 true that $o$ is $P$ to degree 0.3. The philosophical merits of my theory are discussed in detail, and the theory is compared with other extensions and generalisations of fuzzy logic in the literature.

KEY WORDS: blurry sets, degree functions, degrees of truth, fuzzy logic, fuzzy sets, higher-order vagueness, logic, sorites paradox, truth, type *n* fuzzy logic, vagueness

## 1. INTRODUCTION

In this paper I present a new theory of vagueness. By a 'theory of vagueness' I mean a semantics or model theory for a formal language, which is intended to explain and illuminate the relationship between ordinary vague language and the world. At the present stage in the debate about vagueness – at which there is a large number of competing theories in the literature – it is important that one properly motivate any new theory. This is, indeed, such an important task that it ended up taking me a whole paper to do it properly (Smith, 2003). Thus I shall here present only a brief summary of the claims which serve as the background to, and motivation for, the present paper (for detailed presentations of, and arguments for, these claims, see the paper just mentioned, and (Smith, 2001)):

(1) The vagueness literature contains no adequate definition or characterisation of vagueness. We can remedy this situation, and obtain a uniquely useful and attractive way of thinking about vagueness, if we characterise

*Journal of Philosophical Logic* **33:** 165–235, 2004.

vagueness as follows. A predicate '$F$' is vague if and only if it satisfies the following condition (for any objects $a$ and $b$):

*Closeness* If $a$ and $b$ are very similar in $F$-relevant respects, then '$Fa$' and '$Fb$' are very similar in respect of truth.

The basic idea here is that if $a$ and $b$ are very similar in the respects that determine whether an object possesses the property $F$, then they are very similar in respect of possession of $F$ itself. By way of illustration of this characterisation, consider the two predicates 'is at least six feet in height' and 'is tall'. The former predicate is precise and the latter is vague. Suppose that Bill is exactly six feet tall, and Ben is just under six feet tall – say one zeptometre under. Bill and Ben are very close in respects relevant to whether a thing is at least six feet in height; yet 'Bill is at least six feet in height' is true, while 'Ben is at least six feet in height' is false; hence these two sentences are *not* very similar in respect of truth. Closeness is violated here – and this seems right. Now consider the vague predicate 'is tall'. Suppose you take two persons who are not very similar in respects relevant to whether a thing is tall: for example Bob and Bill, who differ in height by two feet. Might it be the case that the claims that Bob is tall and that Bill is tall are not similar in respect of truth? Certainly it might: a significant difference in height can make a significant difference to whether a person is tall. What if Bob and Bill are very close in height: for example if they differ by less than one millimetre? Could it be that the claims that Bob is tall and that Bill is tall are not very similar in respect of truth? Intuitively not: an insignificant difference in height cannot make a significant difference to whether a person is tall. The smaller the difference in height between Bob and Bill, the stronger the intuition: if Bill is only a nanometre shorter than Bill (let alone a picometre, femtometre, attometre, zeptometre or yoctometre shorter), then the claims that Bob is tall and that Bill is tall must, it seems, be *very* similar in respect of truth.

(2) In order to accommodate vagueness as characterised in terms of Closeness, a theory of vagueness must countenance *degrees of truth*.

(3) Almost all of the objections in the literature against theories of vagueness based on fuzzy logic and fuzzy set theory carry no weight. However two objections do carry weight; and indeed one of them is decisive.

The first problem concerns the linear ordering of the fuzzy truth values. Intuitively, given any two vague statements, it need not be the case that one is strictly more true than the other, or else that they have exactly the same degree of truth. The fuzzy view cannot accommodate this idea: on this view, degrees of truth are represented by real numbers in the closed interval [0, 1], and they inherit their truth-ordering from the usual ordering of the reals (which is linear).

The second (and decisive) problem is as follows. There is nothing wrong with representing (for example) Bob's degree of baldness as an element of [0, 1] – *as long as the representation is understood to be merely approximate*. Intuitively, it is *not* correct to say that there is one unique element of [0, 1] that correctly represents Bob's degree of baldness, with all other choices being incorrect. Rather, there are better and worse choices, but none is uniquely correct. Hence, if you say that 'Bob is bald' is true to degree $x$, for some $x$ in [0, 1], it will not in general be the case that your statement is true *simpliciter* or false *simpliciter*: rather, it will be more or less true, according to whether $x$ is a better or worse approximation to Bob's degree of baldness. Now the fuzzy view cannot accommodate this idea: it assigns a unique element of [0, 1] to each sentence, and it is true *simpliciter* that this element is the truth value of that sentence, and false *simpliciter* that any other element is the truth value of that sentence. It is fairly standard to refer to this as the problem of *higher-order vagueness*.[1]

It is important to be clear as to what the higher-order vagueness problem is, and what it is not. There are only two options for a semantics of vagueness: a semantics which *exploits* the base phenomenon of vagueness by presenting its account of vague language *in* vague language, or a semantics which *illuminates* the base phenomenon by giving us a clear understanding – in non-vague terms – of what vagueness involves. There are those who think that there is something wrong with the very idea of assigning a particular semantic status to a vague sentence (whether this status be the possession of a unique classical truth value, a unique fuzzy truth value, a unique set of admissible interpretations, or whatever). These people think that no non-vague theory of vagueness can be correct. But these people are mistaken. We should not abandon the search for a non-vague theory of vagueness. There is no problem with the very idea of associating each vague sentence with one unique truth value. Rather, the higher-order vagueness problem specifically concerns the *fuzzy* truth values: the objection is that we do not get an accurate model of vague language by associating vague predicates with fuzzy sets, and vague sentences with fuzzy truth values.

Consider an analogy. If one repeatedly measures the length of a piece of coastline, using a shorter measuring stick each time, the measured length does not approach a limit. No ordinary curve of Euclidean geometry has this property, and thus it was thought by Mandelbrot that coastlines are not ordinary Euclidean curves. Now the lesson was not that we should not assign *any* particular object to a land-mass as its boundary – it was that we should not assign any object as simple as an ordinary Euclidean curve.

Instead, we should say that coastlines are *fractals*. The problem was not in the very idea of assigning a single curve to a land-mass as its boundary – it was in the assignment of a curve of too simple a sort. Assigning a single fractal is perfectly all right.[2]

Similarly in the case of vagueness: there is *nothing* inherently wrong with the idea of a vague sentence having a unique truth value; there *is* something wrong with the idea that these truth values may be thought of as points between 0 and 1 on the real line. The problems mentioned above are problems with the fuzzy theorist's specific formal proposal for capturing the intuitive idea that properties (including the property of truth) come in degrees. We have no good reason to reject the intuitive idea itself, nor to think that this idea cannot be captured in some *other* way.

(4) There is no way of avoiding these problems with the fuzzy view while retaining [0, 1] as our set of truth values. Specifically, three proposals along these lines fail: (fuzzy) epistemicism; the idea of a hierarchy of fuzzy metalanguages; and the idea of truth being measured on an ordinal (as opposed to absolute) scale.

The upshot of (1)–(4) is as follows. We need degrees of truth, but the real interval [0, 1] cannot play the role of these degrees. Thus we need to find a *different* set of degrees of truth: a set which is such that we *can* feel happy with the idea that each vague sentence is mapped to a unique member of this set. What we need is a non-fuzzy degree-theoretic treatment of vagueness: a treatment which countenances degrees of truth, but which does not identify these with the real numbers between 0 and 1 inclusive. The purpose of the present paper is to provide such a treatment of vagueness.

The abstract picture is as follows. A formal model of the intuitive idea of property-possession to an intermediate degree models properties as sets, where these are functions which assign degrees of membership to objects.[3] Then, if Bob is assigned a certain degree of baldness by the set of bald things, the sentence 'Bob is bald' is assigned that same degree as its truth value – that is, the degree of truth of 'Bob is bald' is the same as Bob's degree of baldness. The fuzzy picture and the picture to be presented in this paper are the same at this abstract level; where they differ is over the formalisation of the intuitive notion of a degree. The fuzzy theorist models properties as fuzzy sets and degrees as numbers in [0, 1]. I model properties as what I call *blurry sets* and degrees as what I call *degree functions*.

The rough idea behind my view is that when faced with a vague statement, such as 'Bob is bald', we may say that it is $x$ true, where $x$ is some real number between 0 and 1, but this will just be a *first approximation* to the actual truth value of the vague statement. For any statement $S_1$ to the effect that ‘Bob is bald’ is $x$ true, there will be a *further* statement $S_2$ which tells us how true $S_1$ is, and so on and on. Thus, while we may say:

(1) ‘Bob is bald’ is 0.6 true

we can then go on to state:

(2) Claim (1) is only 0.7 true

and then further:

(3) Claim (2) is itself only 0.5 true

and so on. Thus we have a hierarchy of statements, none of which tells us the *full and final* story of the degree of truth of ‘Bob is bald’. Now what is really distinctive about the approach in this paper is that the hierarchy is *not* implemented by giving a semantics for vague language which assigns vague sentences real numbers as truth values, and then saying that the metalanguage in which these assignments were made is itself subject to a semantics of the same sort. I regard this approach as both misguided and unworkable (Smith, 2003, §2.5). Rather, the truth values of the system are *not* real numbers: they are degree functions, and the hierarchical structure alluded to above is embedded *inside* each degree function. Each vague sentence is assigned a *unique* degree function as its truth value, and these assignments can be described in a metalanguage whose semantics is *classical*. Thus, instead of a hierarchy of assignments of simple truth values, we have a single assignment of a complex truth value which has an internal hierarchical structure. This approach enables me to avoid the higher-order vagueness problem for the fuzzy view *without* encountering the (to my mind devastating) problems engendered by having a non-classical metalanguage.

The first step in presenting the new theory of vagueness will be to examine these degree functions (or *DF*’s for short): these are the membership/truth values of the system. With an understanding of what they are (Sections 2–4), and of the algebraic properties of the set of all them (Section 5), in hand, we can move on to a development of the theory of blurry sets (Section 6), and from there to a model theory[4] which employs blurry sets and the algebra of degree functions where classical model theory employs ordinary ‘crisp’ sets and the Boolean algebra of classical truth values (Section 7). This model theory is the core of my theory of vagueness. In Section 8 I define a relation of logical consequence with reference to the models introduced in Section 7, and prove that this relation is identical to the classical consequence relation: thus while the *semantics* for vagueness presented here is non-classical, it yields a classical *logic* of vagueness. In Section 9 I make good on the informal idea which I use to motivate my view: the idea that while, for example, we may say that ‘Bob is bald’ is 0.6 true, we can then go on to state that our previous claim was only 0.7 true, and then further that this most recent statement was itself only 0.5 true, and so on. I do this by showing that we can consistently introduce infinitely many truth predicates into the formal language discussed in the paper, and that these predicates will behave in the desired ways. On the basis of this section, I go on in Section 10 to respond to an argument of Timothy Williamson which is designed to show that anyone who denies bivalence (as I do in this paper) is committed to asserting a contradiction; I show where Williamson’s argument goes wrong. In Section 11 I show that the view presented here really does make room for higher-order vagueness, and thus solves the main problem with the fuzzy view which it was designed to solve; and in Section 12 I show how my view deals with sorites paradoxes. In Section 13 I show how my view can be extended to accommodate multiple admissible interpretations (i.e. semantic vagueness or ambiguity). Finally in Section 14 I compare my view to various other extensions and generalisations of fuzzy logic in the literature. I show that my theory is different from existing views both in its formal details, and – more importantly – in its conceptual foundations.

## 2. Degree Functions

Two ideas emerge naturally from the higher-order vagueness problem for the fuzzy account, and the motivation behind my account is to accommodate these ideas.

The first idea is that while it would be fine to give a fuzzy degree (i.e. a real number between 0 and 1) as a rough approximation to Bob’s degree of baldness, or to the degree of truth of some vague statement, a fuzzy degree cannot tell the *full* story of Bob’s degree of baldness (or any other vague matter), hence nor can such a number tell the full story of the degree of truth of ‘Bob is bald’. For any statement $S_1$ to the effect that ‘Bob is bald’ is $x$ true, for $x$ in [0, 1], it seems there should be a *further* statement $S_2$ which tells us how true $S_1$ is, and so on (perhaps for ever, or perhaps not). We should have a hierarchy of statements: a first-order statement assigning a fuzzy degree of baldness to Bob; second-order statements assigning fuzzy degrees of truth to each possible first-order statement; and so on.

The second idea is that there is *something* right about modelling vagueness in terms of the closed interval [0, 1], but the fuzzy approach develops

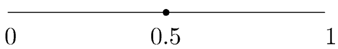


*Figure 1.* Bob's degree of baldness – first approximation.

this idea in too simple-minded a way. Bob's degree of baldness is not well modelled by a *point* in this interval. Rather, it would be better to model his degree of baldness by a blurry or cloudy region stretching somewhere between 0 and 1, with a higher density in some areas than in others. Bob's degree of baldness cannot be said to be 0.6, or any other number, exactly – but, the idea goes, it can be said to cluster around 0.6, trailing off on each side. The idea is that Bob's degree of baldness may be in between complete (i.e. 1) and non-existent (i.e. 0), *without* being at any *particular* point in between. There is no unique point which provides the correct answer to the question 'What is Bob's degree of baldness?', but some points are better choices than others.

My model of the notion of a degree is based on the idea that a degree – say, Bob's degree of baldness – can be specified by giving a sequence of better and better *approximations*, each approximation taking the form of a *fuzzy* degree. It will thus be important to distinguish three things: the intuitive notion of a degree; my formalisation of this notion, which is a *DF* (degree function); and a fuzzy degree, or real number in [0, 1], which is the fuzzy theorists' (failed) attempt at a formalisation of the intuitive notion. In my account, degrees are *modelled* or *represented* by *DF*'s, and are *approximated* by fuzzy degrees.

The picture is as follows. I can say that Bob is, say, 0.5 bald (or bald to degree 0.5), but this will only be a first approximation to the full story of his degree of baldness (Figure 1). If I am to give more detail about Bob's degree of baldness, I will need to tell you the *degree* to which Bob is 0.5 bald, and also the degree to which he is $x$ bald, for other values of $x$ in [0, 1]. Just as I initially approximated Bob's degree of baldness with a fuzzy degree, I shall initially approximate these further degrees in the same way. Thus, at the second stage of my account of Bob's degree of baldness I give you one number in [0, 1] for *each* number in [0, 1]. We may think of all these numbers as plotting out a curve: the graph of a function from [0, 1] to [0, 1], as in Figure 2.[5] The idea here is that the degree to which Bob is bald to degree 0.5 is 0.7 and the degree to which Bob is bald to degree 0.3 is 0.5. In terms of degrees of truth, the idea is that the sentence ' 'Bob is bald' is 0.5 true' is 0.7 true, while the sentence ' 'Bob is bald' is 0.3 true' is 0.5 true. Thus what we are given at this second stage of approximation is an assessment of the different possible first approximations of Bob's degree of baldness: the higher the number assigned to $x$, the better a first approximation $x$ is – and the important point is that it will not in general

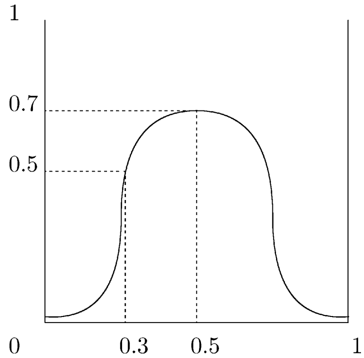


*Figure 2.* Bob's degree of baldness – second approximation.

be the case that one number $x$ in [0, 1] is assigned 1 while all the others are assigned 0.

The second idea mentioned above was that Bob's degree of baldness should be represented as a blurry region stretching somewhere between 0 and 1, rather than as a single point. This idea is accommodated by regarding the curve just described as the graph of a density function: the higher the curve is over point $x$, the greater the density of Bob's degree of baldness at that point. Bob's degree of baldness will not, in general, be located all at one point; rather, how much of it is located in any sub-interval $[a, b]$ of [0, 1] is given by the area under the curve between $a$ and $b$. Letting $f(x) : [0, 1] \to [0, 1]$ be the function of which the curve is the graph, this area is $\int_a^b f(x)\,\mathrm{d}x$. Thus we will want to suppose that the function $f(x)$ is integrable. We will also want to suppose that it is normalised: that is, the area under the entire curve is 1, i.e. $\int_0^1 f(x)\,\mathrm{d}x = 1$. The idea here is that Bob's degree of baldness is entirely confined within the interval [0, 1]: it may not all be at one point therein, but none of it is anywhere outside the interval.

Now the only functions $f(x) : [0, 1] \to [0, 1]$ which bound a region of area 1 are the constant function which takes the value 1 for every argument (Figure 3), and functions which differ from it at no more than countably many points. In order to allow for a greater variety of density functions – while still restricting ourselves to functions from [0, 1] to [0, 1], rather than from [0, 1] to the set **P** of non-negative reals (this restriction will be important later) – we may simply regard the function from [0, 1] to [0, 1] of which the curve described above is the graph as an *encoded* density function, as follows. There are as many numbers in **P** as there are in [0, 1), and there are many ways of specifying an isomorphism between these two sets. Any homeomorphism (i.e. *continuous* isomorphism) $\mathfrak{f} : [0, 1) \to \mathbf{P}$ will serve our purposes – for example, the canonical

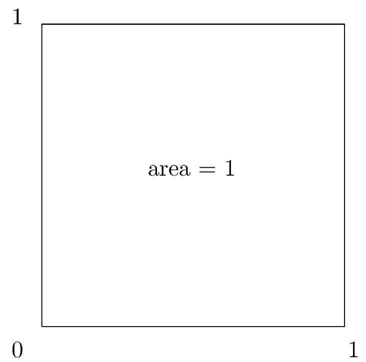


*Figure 3.* A normalised density function [0, 1] → [0, 1].

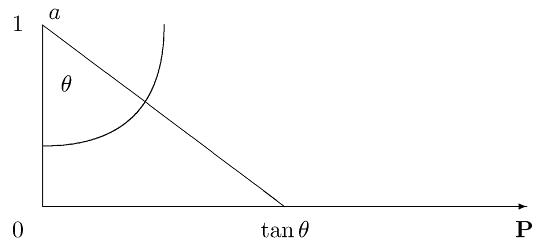


*Figure 4.* The canonical homeomorphism [0, 1) → **P**.

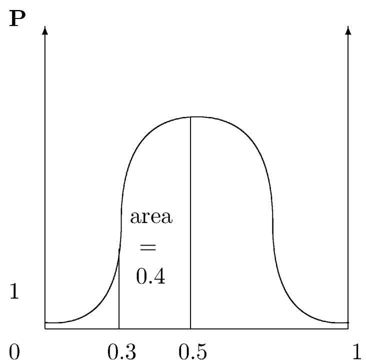


*Figure 5.* The density function encoded by the second approximation.

homeomorphism, which takes $x \in [0, 1)$ to $\tan(90x)^\circ$ (Figure 4).[6] Now, given $\mathfrak{f}$, which translates talk of [0, 1) into talk of **P**, we may regard our function $f(x) : [0, 1] \to [0, 1]$ as an encoded version of the function $(\mathfrak{f} \circ f)(x) : [0, 1] \to \mathbf{P}$, and we may suppose that the *latter* is a normalised density function over [0, 1]. Figure 5 shows the function $\mathfrak{f} \circ f$, where $f$ is the function shown in Figure 2. In Figure 5, the area under the entire curve is 1. Thus, if we ask how much of Bob's degree of baldness is located between 0 and 1, the answer will be 100% – but unlike in the fuzzy account, where Bob's degree of baldness is identified with a single point, in the present account it might be that 90% of Bob's degree of baldness is between 0.2 and 0.8, 50% between 0.3 and 0.7, 10% between 0.48 and 0.52, and so on. For example, if the area under the curve between 0.3 and 0.5 is 0.4, then 40% of Bob's degree of baldness is distributed between 0.3 and 0.5.

The story so far: In describing Bob's degree of baldness, we may begin with an approximation: a fuzzy degree, or real number between 0 and 1. We may then proceed to a more detailed approximation, by associating one number in [0, 1] with each number in [0, 1] – that is, by giving a function $f(x) : [0, 1] \rightarrow [0, 1]$. The first intuitive idea is captured as follows: suppose the number associated with 0.3 is 0.4; then the idea is that it is 0.4 true that Bob's degree of baldness is 0.3. The second intuitive idea is captured as follows: $f$ encodes a density function $\mathfrak{f} \circ f$, which tells us the density of Bob's degree of baldness at every point between 0 and 1, and hence allows us to calculate the amount of Bob's degree of baldness located in any sub-region $[a, b]$ of [0, 1]. Finally, I impose one additional requirement: the initial approximation to Bob's degree of baldness is the same as the centre of mass of the density function specified at the second level of approximation. This ensures that the second level of approximation really can be thought of as an *improvement* on the initial approximation: rather than being totally unrelated to the initial approximation, the second level of approximation simply provides a further level of detail which was ignored at the initial level of approximation.

A word of warning is required at this point concerning the interpretation of the numbers given at the second stage of approximation. These numbers should be thought of as *densities*, not as *masses*. Thus, nothing has gone wrong if I tell you that it is 0.3 true that Bob is bald to degree 0.2, 0.4 true that Bob is bald to degree 0.3, 0.5 true that Bob is bald to degree 0.4, 0.4 true that Bob is bald to degree 0.5, 0.3 true that Bob is bald to degree 0.6, and so on – that is, there is no requirement that the numbers assigned to the elements of [0, 1] at the second level of approximation *sum to* 1. Rather, what must equal 1 is the area under the graph of the density function that all these numbers encode. The area under any segment of this curve (unlike the height of the curve at any point) should be thought of as a mass, not as a density. This is why I reserved the use of percentages for talk of such areas. It would sound very odd to say that it is 50% true that Bob's degree of baldness is between 0 and 0.3, and 60% true that it is between 0.3 and 0.5: if we divide the curve into disjoint segments, then the areas under these segments should sum to 1.[7] Thus it does not matter that the encoding function $\mathfrak{f}$ is arbitrarily chosen. For different choices of $\mathfrak{f}$, the numbers assigned at the second level of approximation will be different: for one choice it might be 0.4 true that Bob is 0.3 bald, for another choice it might be 0.3 true that Bob is 0.3 bald. But this does not matter, because these numbers were never particularly meaningful in isolation anyway: what matters is the density function which, *taken together*, they encode. It is significant that 20%, rather than 30%, of Bob's degree of baldness is between 0.2 and 0.3; it is not significant that, considered in isolation, the value assigned to 0.2 at the second level of approximation is 0.2, rather than 0.3. Thus, if you ask to what degree Bob is bald, and I say 0.3, and then you ask how true my previous statement was, and I say 0.4, this latter number does not tell you very much by itself – but given the answers to all questions of the form 'How true is it to say that Bob is bald to degree $x$', for $x$ in [0, 1], you do indeed have significantly more information about Bob's degree of baldness than you had when I simply said he was bald to degree 0.3: you know that Bob's degree of baldness is distributed thus and so between 0 and 1, with the centre of mass of the distribution being at 0.3, and for any subregion $[a, b]$ of [0, 1], you can determine what proportion of Bob's degree of baldness is in that region.[8]

In describing Bob's degree of baldness, we have now seen two levels of approximation. This is not the end of the story. Each of the fuzzy degrees that I gave at the second level of approximation is itself a first approximation to the degree to which Bob is bald to degree $x$, for each $x$ in [0, 1]. Thus we may progress to a third stage of approximation, at which we fill out the approximations given at the second stage. No new ideas are involved in this further step: because we have restricted ourselves to functions from [0, 1] to [0, 1], we can simply iterate the picture presented above. Just as the initial approximation turned out, at the next level of approximation, to be the centre of mass of a density function, so too each of the fuzzy degrees which together specified that function turns out, at the next level of approximation, to be the centre of mass of a further density function: one such function for each such fuzzy degree. In Figure 6, one such further density function is shown: the one associated with the point (0.5, 0.7) on the curve given at the second level of approximation. This is just for ease of illustration: it is to be understood that at the third level of approximation, *each* point on the second-level curve is associated with a further density function, of which that point is the centre of mass. The idea is that (for example) the degree to which (the degree to which Bob is bald to degree 0.5 is 0.7) is 0.8 – or in terms of degrees of truth, the idea is that the sentence ' ' 'Bob is bald' is 0.5 true' is 0.7 true' is 0.8 true. Thus what we are given at this third stage of approximation is an assessment of the different possible second approximations of Bob's degree of baldness.

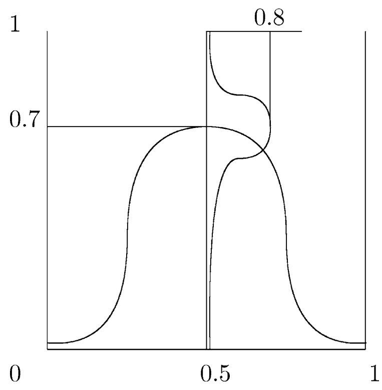


*Figure 6.* Bob's degree of baldness – third approximation (part view).

At the first stage we were given a number, and at the second stage each possible first-stage number was given a number. At the third stage, each possible second-stage number that could have been given to some first-stage number is itself given a number, indicating how good a choice it would have been at the second stage – and as before, it will not in general be the case that for any fixed first-stage number, one possible second-stage number $x$ in [0, 1] is assigned 1 at the third stage, while all the others are assigned 0.

Now I still haven't finished telling you about Bob's degree of baldness. The latest level of approximation may be filled out by a further level of approximation, which may be filled out by a further level, and so on and on. Bob's degree of baldness (or any other degree) is thus thought of as an infinite hierarchy of better and better approximations, where these approximations consist of fuzzy degrees, or numbers in [0, 1]; the move from one level of approximation to the next consists in the replacement of each number $x$ in [0, 1] by a density function over [0, 1], with centre of mass at $x$. At any level, what I need to do to *finish* the story is tell you the *degree* to which ... (the degree to which Bob's degree of baldness is $x$ is $y$) ... is $z$, but all I actually do is *approximate* these degrees by further fuzzy degrees. However, after $\omega$ stages of approximation, all the approximations have been filled out, and we then have the *full* story of Bob's degree of baldness. One way of picturing such a degree is as a region of varying shades of grey spread between 0 and 1 on the real line. If you focus on any point in this region, you see that what appeared to be a point of a particular shade of grey is in fact just the centre of a further such grey region. The same thing happens if you focus on a point in this further region, and so on. The region is blurry all the way down: no matter how much you increase the magnification, it will not come into sharp focus.

## 3. Formalisation

I have now presented a relatively informal account of how we might model the intuitive notion of a degree, and it is time to make these ideas more precise. Formally, a degree will be represented by a function which I call a degree function, or *DF*. Each *DF* is a function $f : [0,1]^* \to [0,1]$, where $[0,1]^*$ is the set of words on the alphabet [0, 1] (i.e. the set of all finite sequences of elements of [0, 1], including the empty or null sequence). The assignment to the empty sequence $\langle\rangle$ is the initial approximation. The assignments to sequences of length 1 together constitute the second level of approximation: for example, where $f$ is object $x$'s degree of property $X$, if $f(\langle 0.3\rangle) = 0.4$, then it is is 0.4 true that $x$ is $X$ to degree 0.3; the assignments to sequences of length 2 together constitute the third level of approximation: for example if $f(\langle 0.3, 0.4\rangle) = 0.5$, then it is 0.5 true that it is 0.4 true that $x$ is $X$ to degree 0.3; and so on.

Let $F$ be the set of all functions $f : [0,1]^* \to [0,1]$, i.e. $F = [0,1]^{([0,1]^*)}$. Let *DF* be the set of all *DF*'s. Every *DF* is in $F$, but not vice versa. The members of *DF* are of three distinct types.

*Type I (Basic).* The first type of *DF* corresponds to the idea of a degree developed in the previous section. For any $f \in F$, consider the values assigned by $f$ to the sequences $\langle a_1, \ldots, a_n, x\rangle$, for fixed $a_1, \ldots, a_n$ and variable $x \in [0,1]$. These values determine a function from [0, 1] to [0, 1], which we shall denote $f_{\langle a_1,\ldots,a_n\rangle}$:

$$f_{\langle a_1,\ldots,a_n\rangle}(x) = f(\langle a_1, \ldots, a_n, x\rangle).$$

(In the special case $n = 0$, we have $f_{\langle\rangle}(x) = f(\langle x\rangle)$.) In a Type I *DF* $f$, we require that for every sequence $\langle a_1, \ldots, a_n\rangle$, $\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n\rangle}$ is an integrable function with the property that

$$\int_0^1 (\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n\rangle})(x)\,\mathrm{d}x = 1.$$

This ensures that we may regard each $f_{\langle a_1,\ldots,a_n\rangle}$ as an encoded normalised density function over [0, 1].[9] We also require that

$$f(\langle a_1, \ldots, a_n\rangle) = \int_0^1 x(\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n\rangle})(x)\,\mathrm{d}x.$$

This is the centre of mass requirement discussed above.[10] This requirement ensures that there is a suitable relationship between the successive levels of approximation: if, at the third level of approximation, $f(\langle 0.3, 0.4\rangle) = 0.5$ – that is, it is 0.5 true that it is 0.4 true that $x$ is $X$ to degree 0.3 – then at the fourth level of approximation, $f_{\langle 0.3,0.4\rangle}$ encodes a density function whose centre of mass is 0.5.

There are some situations which we would like to accommodate, which are not captured by any Type I *DF*. Thus we introduce two further types of *DF*.

*Type II*. If Bob is a definite, clear case of a bald man, then to a first approximation, he should be bald to degree 1; to a second approximation, it should be 1 true that he is bald to degree 1 and 0 true that he is bald to any degree other than 1; to a third approximation, it should be 1 true that it is 1 true that he is bald to degree 1, 0 true that it is $x$ true that he is bald to degree 1 for $x \neq 1$, 1 true that it is 0 true that he is bald to any degree other than 1, and so on. Similarly, if Bob is a definite, clear countercase of a bald man, then to a first approximation, he should be bald to degree 0; to a second approximation, it should be 1 true that he is bald to degree 0 and 0 true that he is bald to any degree other than 0; to a third approximation, it should be 1 true that it is 1 true that he is bald to degree 0, 0 true that it is $x$ true that he is bald to degree 0 for $x \neq 1$, 1 true that it is 0 true that he is bald to any degree other than 0, and so on. Thus we introduce two *DF*'s **T** and **F**:[11]

- $\mathbf{T}(\langle\rangle) = 1$.
- If $\mathbf{T}(\langle x_1, \ldots, x_n\rangle) = k$, then $\mathbf{T}(\langle x_1, \ldots, x_n, k\rangle) = 1$ and $\forall j \neq k$, $\mathbf{T}(\langle x_1, \ldots, x_n, j\rangle) = 0$.
- $\mathbf{F}(\langle\rangle) = 0$.
- If $\mathbf{F}(\langle x_1, \ldots, x_n\rangle) = k$, then $\mathbf{F}(\langle x_1, \ldots, x_n, k\rangle) = 1$ and $\forall j \neq k$, $\mathbf{F}(\langle x_1, \ldots, x_n, j\rangle) = 0$.

These are the analogues of the classical The True and The False: if $x$ is $X$ to degree **T**, then $x$ is *utterly* $X$, and if $x$ is $X$ to degree **F**, then $x$ is not $X$ *at all*.

*Type III*. We want to allow for the following sort of possibility: to a first approximation Bob is bald to degree 0.5; to a second approximation it is 0.6 true that Bob is bald to degree 0.5, 0.5 true that Bob is bald to degree 0.6, and so on; *but at the third level of approximation the second level approximations are simply ratified*: it is 1 true that it is 0.6 true that Bob is bald to degree 0.5, and 0 true that it is $x$ true that Bob is bald to degree 0.5 for $x \neq 0.6$; it is 1 true that it is 0.5 true that Bob is bald to degree 0.6, and 0 true that it is $x$ true that Bob is bald to degree 0.6 for $x \neq 0.5$, and so on. Likewise, the third level approximations are simply ratified at the fourth level, and so on. (In general, we want to allow for the possibility that at the $n$th level of approximation, for *any* $n \geq 2$, and then at all higher levels of approximation as well, the lower-level approximations are simply ratified.) To accommodate this sort of possibility, we introduce a family $\{f_n\}$ of Type III *DF*'s (one for each non-negative integer $n$) for *each* Type I *DF*. The idea is that where $f$ is a Type I *DF*, $f_n$ makes the same assignments as $f$ to sequences of lengths up to $n$, and then for longer sequences – that is, at subsequent levels of approximation – $f_n$ simply ratifies the assignments it has made at lower levels. Thus:

– For $i \leq n$, $f_n(\langle x_1, \ldots, x_i \rangle) = f(\langle x_1, \ldots, x_i \rangle)$.
– For $i \geq n$, if $f_n(\langle x_1, \ldots, x_i \rangle) = k$, then $f_n(\langle x_1, \ldots, x_i, k \rangle) = 1$ and $\forall j \neq k$, $f_n(\langle x_1, \ldots, x_i, j \rangle) = 0$.

## 4. RESTRICTIONS

A basic set of *DF*'s would contain all and only the members of $F$ of Types I, II and III just specified. I have been very liberal, however, and admitted far more members of $F$ than I needed to in order to accommodate the intuitive picture of a degree sketched out in Section 2. In fact I have admitted $2^{\mathfrak{c}}$ members of $F$.[12] I shall consider two ways in which we might further restrict membership in *DF*. Both restrictions focus on Type I *DF*'s: the definitions of Types II and III are left unchanged. However, it is to be understood that because Type III is defined parasitically on Type I (we define a family of Type III *DF*'s for each Type I *DF*), as the membership of Type I changes, so does the membership of Type III. (The membership of Type II remains constant.)

*Type I (Continuous)*. We retain the conditions on Type I (Basic) *DF*'s, and add two more. First, where we required that for every sequence $\langle a_1, \ldots, a_n \rangle$, $\mathfrak{f} \circ f_{\langle a_1, \ldots, a_n \rangle}$ must be an integrable function, we now furthermore require that it must be a *continuous* function. Second, for each $f \in F$ and each positive integer $n$, let the $n$-place function $f^n$ be defined as follows:

$$f^n(x_1, \ldots, x_n) = f(\langle x_1, \ldots, x_n \rangle).$$

Then we require that for a Type II (Continuous) *DF* $f$, $f^n$ is continuous for every positive integer $n$.

Intuitively, there is no object $x$ and property $X$ such that the degree to which $x$ possesses $X$ corresponds to a member of $F$ ruled out by these continuity requirements – that is, the excluded *DF*'s do not meet any intuitive need (unlike, for example, **T** and **F**, which were therefore added to the original Type I (Basic) *DF*'s). Furthermore, these requirements render the set of *DF*'s more manageable by reducing its cardinality to $\mathfrak{c}$.[13]

Just as the properties of classical set theory and logic follow in a natural way from the properties of the Boolean algebra of classical truth values (and the properties of fuzzy set theory and logic follow in a natural way from the properties of the Kleene algebra of fuzzy truth values), so the properties of blurry set theory and logic will follow in a natural way from the properties of the algebra of degree functions – once we determine what that algebra is. We have a set *DF* of degree functions, and the question will be: are there any natural operations $\vee$, $\wedge$ and $'$ on this set which can be used to define union, intersection and complementation for blurry sets, and the truth conditions of disjunctive (and existentially quantified), conjunctive (and universally quantified) and negated sentences? One attractive idea is that if we take notice only of first approximations – that is, if we ignore everything about a *DF* except what it assigns to the empty sequence – then things should work just as they do in ordinary fuzzy logic. That is, where $f$ and $g$ are *DF*'s:

$$f'(\langle\rangle) = 1 - f(\langle\rangle)$$
$$(f \vee g)(\langle\rangle) = \max\{f(\langle\rangle), g(\langle\rangle)\}$$
$$(f \wedge g)(\langle\rangle) = \min\{f(\langle\rangle), g(\langle\rangle)\}.$$

But even with this restriction in place, there is still a huge range of possible operations on *DF*: for given two *DF*'s $f$ and $g$, all we have fixed about $f \vee g$, $f \wedge g$ and $f'$ is what they assign to the empty sequence, and this leaves an enormous amount open. Given $f$ and $g$, what should $f \vee g$, $f \wedge g$ and $f'$ look like, above the first level of approximation? Various ideas can be explored here, but the problem is that there does not seem to be a unique best idea: for any proposed way of fixing the upper levels, there seem to be plenty of other ways that are just as deserving of the titles 'disjunction'/'union', 'conjunction'/'intersection', and 'negation'/'complementation'. Thus there are two options which we might consider. One approach is to work at a relatively abstract level, allowing at each stage for the full range of possible algebraic operations on *DF*. The second approach is to place further constraints on the membership of *DF* itself, in such a way that there is only one natural choice of operations on the reduced set of degree functions. The first approach proves to be more tractable than one might imagine, but it is really less interesting than the second approach, which I shall therefore pursue now.[14]

As things currently stand, any (encoded) continuous normalised density function is admissible as the second level of approximation to $x$'s degree of $X$. This leaves open an enormous range of possibilities, most of which have no intuitive content; for example, see Figure 7. Of course we can interpret this picture, in accordance with the explanations given earlier in this paper. The point is not that this (encoded) density function is *meaningless*, rather that it is *useless*: we cannot think of a sentence and a situation such

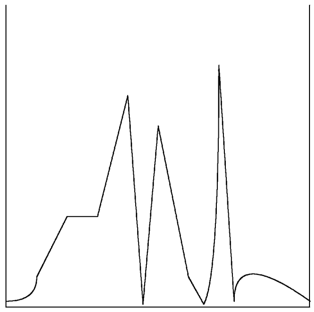


*Figure 7.* A density function with no intuitive content.

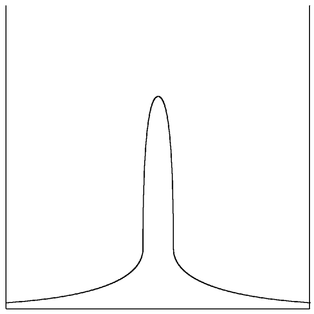


*Figure 8.* Bob's degree of baldness (second approximation).

that in that situation, the sentence would (to a second approximation) have this truth value.

One distinction amongst density functions that *does* have intuitive content is that between 'tall and skinny' and 'short and fat'. If Bob's degree of baldness is (to a second level of approximation) as in Figure 8 and Bill's degree of baldness is (to a second level of approximation) as in Figure 9, then while each man's degree of baldness is evenly distributed around 0.5, Bob's is more localised. The following terminology is natural. Where $f$ is the *DF* of sentence $S$, $S$ is first-order vague if $f(\langle\rangle) \in (0, 1)$, that is, if (to a first approximation) $S$ has an intermediate degree of truth; $S$ is second-order vague if $f(\langle x\rangle) \in (0, 1)$, for some $x \in [0, 1]$, that is if, at the second level of approximation, some sentence of the form '$S$ is $x$ true' has an intermediate degree of truth; $S$ is third-order vague if $f(\langle x, y\rangle) \in (0, 1)$, for some $x, y \in [0, 1]$, that is if, at the third level of approximation, some sentence of the form ''$S$ is $x$ true' is $y$ true' has an intermediate degree of truth; and so on.[15] Furthermore, it is natural to distinguish *grades* of higher-order (sentence) vagueness: in the case of Bob and Bill, both 'Bob

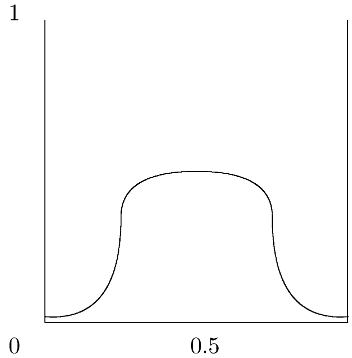


*Figure 9.* Bill's degree of baldness (second approximation).

is bald' and 'Bill is bald' are first-order vague and second-order vague – and it seems natural to say that the latter is *more* second-order vague than the former.

We thus want to be able to make at least the following distinctions amongst *DF*'s. First, two *DF*'s $f$ and $g$ may be distinguished by the assignments they make to the empty sequence. Second, they may be distinguished according to whether the density functions encoded by the assignments they make to sequences of length 1 are more or less 'tall and skinny' or 'short and fat'. Now consider sequences of length 2, and distinguish two different scenarios. In the first, for every $x$ and $y$, $f_{\langle x \rangle}$ and $f_{\langle y \rangle}$ have the same degree of spread (i.e. of 'shortness and fatness'/'tallness and skinniness'). In this case, the measure of spread[16] serves as a measure of third-order vagueness. Thus this scenario has intuitive content. In the second case, there are $x$ and $y$ such that $f_{\langle x \rangle}$ is relatively tall and skinny while $f_{\langle y \rangle}$ is relatively short and fat. Although, as above, we can interpret this case in light of the explanations I gave of degree functions, it does not seem to be a case that is intuitively useful – and thus is not a scenario for which we need to allow. Similarly, when it comes to sequences of length 3, we want to allow for the cases in which, for every $x$, $y$, $z$, $w$, $f_{\langle x,y \rangle}$ and $f_{\langle z,w \rangle}$ have the same degree of spread, but we do not need to allow for the cases in which, for some $x$, $y$, $z$, $w$, $f_{\langle x,y \rangle}$ is relatively tall and skinny while $f_{\langle z,w \rangle}$ is relatively short and fat – and so on for longer sequences.

Do we need to make any further distinctions amongst *DF*'s than the ones just mentioned? I cannot see that we do – but I can think of a reason why someone might mistakenly think so. Suppose that Bob does not have very much hair, but that what he has evenly covers his scalp. On the one hand it seems that Bob's degree of baldness is quite high (low hair count), while on the other hand it seems quite low (even scalp coverage). Thus one *might* think that to a second approximation, Bob's degree of baldness

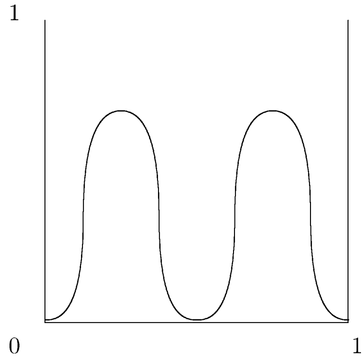


*Figure 10.* A density function with two local maxima.

should be as in Figure 10. The idea would be that to a first approximation, we ‘average out’ all the relevant factors (hair count and scalp coverage), coming to the conclusion that Bob is 0.5 bald. At the second level of approximation, however, we are more subtle: we say that there is a good deal of truth in the claim that Bob is 0.25 bald or thereabouts (because he has even scalp coverage) and also a good deal of truth in the claim that Bob is 0.75 bald or thereabouts (because he has a low hair count), but there is not a good deal of truth in the claim that Bob is 0.5 bald or thereabouts: saying that Bob is 0.5 bald is the right thing to do at the first, roughest level of approximation, but not so at this more subtle level.

There are two possible thoughts behind this suggestion. First, the idea might be that the case in question is one of ambiguity or semantic vagueness. The situation is not that Bob has one particular property, *baldness*; rather, there are two senses of ‘bald’: the ‘few hairs’ sense and the ‘low scalp coverage’ sense. In one sense Bob is fairly bald (that is, he possesses to a fairly high degree the property picked out by this sense of ‘bald’), in the other sense fairly non-bald (that is, he possesses to a fairly low degree the property picked out by this other sense of ‘bald’). If this is the idea, however, then it is not properly captured in the way just indicated. Degree functions are meant to help us model *worldly* vagueness: if ‘Bob is bald’ has a particular *DF* $f$ as its degree of truth, then there is a particular property (blurry set) picked out by ‘is bald’, and the degree to which Bob possesses this property is $f$. Semantic vagueness *can* be accommodated within an extension of the framework outlined in this paper – as I shall explain in Section 13 – but it is *not* properly accommodated in the way indicated above.[17]

Second, the idea might be that ‘bald’ does indeed pick out a unique property, but that this property is multi-dimensional: whether or not (or to what degree) an object possesses it is determined by a number of factors –

in this case hair count and scalp coverage. Now this may very well be so – indeed it seems to me to be a better analysis of 'bald' than the one just considered, according to which 'bald' is ambiguous. Nevertheless, there is no good reason to treat multi-dimensional vagueness within the framework presented in this paper in the way suggested above. A better idea is as follows. Suppose that $x$'s degree of $X$ is a function of $x$'s position on various scales. Distinguish the case in which $x$ is halfway along each scale, from the case in which $x$ is high on some scales and low on others. In both cases, it seems natural to say that $x$'s degree of $X$ is about 0.5 (to a first approximation); however it seems that the statement '$x$ is $X$' is more second- (and perhaps higher-) order vague in the latter case than in the former. Thus, in the second case we want the *DF* of the sentence '$x$ is $X$' to encode, at the second level of approximation, a density function which is relatively short and fat. Thus, in order to capture the phenomenon at issue, we do not need to appeal to *DF*'s of the sort pictured in Figure 10: the 'tall and skinny' versus 'short and fat' distinction is enough. The thought that we need the more elaborate *DF*'s seems to me to derive from a conflation of multi-dimensionality and ambiguity, combined with a wrongheaded idea about how to accommodate ambiguity within the semantic framework presented in this paper.

So we want to allow ourselves enough *DF*'s to make the 'tall and skinny' versus 'short and fat' distinction, and no more. A natural thought is to say that for every sequence $\langle a_1, \ldots, a_n \rangle$ (including the empty sequence), $\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n \rangle}$ is a *normal* density function[18] – that is, there are $\mu$ and $\sigma$ such that:

$$(\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n \rangle})(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}.$$

The normal distribution crops up frequently in probability theory: its graph is the familiar *bell curve* (Figure 11). Why, then, is it natural to use a

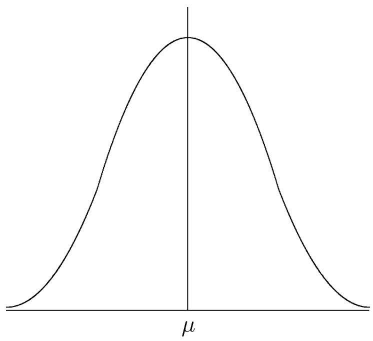


*Figure 11.* The normal distribution.

curve of this shape in the present context, where the interpretation is *not* probabilistic? Well, it should first be said that I am not *deeply* committed to the normal distribution. I need to be able to distinguish amongst density functions according to their centres of mass, and according to how spread out they are ('tall and skinny' versus 'short and fat'), and the normal distribution is a well-known type of function that allows us to make precisely these distinctions. A normal density function is determined by two numbers, $\mu$ and $\sigma$. In the probabilistic interpretation, $\mu$ is the mean and $\sigma$ is the standard deviation; in the present context, $\mu$ is the centre of mass, and $\sigma$ is a measure of spread. This said, however, the normal distribution also seems a natural choice for a further reason. Suppose that Bob is a borderline case for 'bald', and suppose we ask a group of people to rate Bob's degree of baldness on a scale from 0 to 100 (or equivalently, from 0.00 to 1.00), where 0 is what one would assign to Fabio and 100 is what one would assign to Yul Brynner.[19] We would expect the results to be normally distributed. Now the normal distribution often crops up in connection with the distribution of random errors – it was studied by Gauss in this context, and indeed is also known as the Gaussian or *error* curve. In the present case, however, we do *not* think that there is one correct answer to the question of where Bob's degree of baldness belongs on a scale from 0 to 100. Rather, we think that if forced (as in this case) to assign a particular number, there are better and worse options, but none is uniquely correct. Thus it is natural for us to interpret the mean of the distribution as a first approximation to Bob's degree of baldness, and to take the standard deviation as a measure of the second-order vagueness of the claim that Bob is bald – that is, as a measure of how much weight we can put on the initial approximation (say 70), to the exclusion of other possible values (such as 65 and 75). We thus want to take the *shape* of the normal distribution, but *not* also take on board its usual probabilistic connotations.

Now we cannot say *simply* that for all sequences $\langle a_1, \ldots, a_n \rangle$, $\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n \rangle}$ is a normal density function, because while for any such function $f(x)$, $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$, the 'tails' of the normal distribution never reach the horizontal axis (see Figure 11), so for any finite $a$ and $b$, $\int_a^b f(x)\,\mathrm{d}x < 1$, whereas we want $\int_0^1 (\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n \rangle})(x)\,\mathrm{d}x = 1$. Rather than abandon the normal distribution, however, we may simply stipulate that for each $\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n \rangle}$, there is a unique normal density function $N(\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n \rangle})$ associated with it, which it resembles extremely closely: they are not identical, but they are close enough that we may (for almost all purposes) think of the graph of $\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n \rangle}$ as simply being a bell curve. One of the properties of the normal distribution is that 99.99% of observations fall within ±4 standard deviations from the mean. Thus, if a normal density function $f(x)$ has mean $0 < \mu < 1$ and standard deviation

$$0 < \sigma \leq \frac{1}{4}\min\{\mu, 1 - \mu\}$$

then $\int_0^1 f(x)\,\mathrm{d}x > 0.9999$. Hence to be a very close approximation to some function $\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n\rangle}$, a normal density function must have mean and standard deviation within the ranges just specified.[20]

Suppose we have a normal density function $f$ with mean and standard deviation within these ranges. Define $f$'s *precision* $\pi$ as follows:

$$\pi = 1 - \frac{\sigma}{\frac{1}{4}\min\{\mu, 1 - \mu\}}$$

$0 \leq \pi < 1$. If $f$'s standard deviation is as large as it can be, namely equal to $\frac{1}{4}\min\{\mu, 1-\mu\}$, then $f$'s precision is 0; the smaller $f$'s standard deviation, the greater its precision. The idea behind this definition of precision is to make precise the earlier distinction between 'tall and skinny' and 'short and fat': the greater the precision of $N(\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n\rangle})$, the less spread out – the more tall and skinny – is its graph. $\pi$ will be our measure of spread.

We can now state a new definition of Type I *DF*'s.

*Type I (Normal).* In a Type I *DF* $f$, we require that for every sequence $\langle a_1, \ldots, a_n\rangle$, $\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n\rangle}$ is associated with a unique normal density function $N(\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n\rangle})$ which approximates $\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n\rangle}$ extremely closely, and has mean $0 < \mu < 1$ and standard deviation $0 < \sigma \leq \frac{1}{4}\min\{\mu, 1-\mu\}$. We also impose the familiar 'centre of mass' requirement:

$$f(\langle a_1, \ldots, a_n\rangle) = \int_0^1 x(\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n\rangle})(x)\,\mathrm{d}x.$$

Finally, we require that for any sequences $\langle a_1, \ldots, a_i\rangle$ and $\langle b_1, \ldots, b_j\rangle$, if $i = j$ (i.e. the sequences are the same length) then $N(\mathfrak{f} \circ f_{\langle a_1,\ldots,a_i\rangle})$ and $N(\mathfrak{f} \circ f_{\langle b_1,\ldots,b_j\rangle})$ have the same precision.[21]

As before, the membership of Type II remains the same (**T** and **F**), and the membership of Type III changes in line with the change to Type I: for each Type I *DF* $f$ and each non-negative integer $n$ there is a Type III *DF* $f_n$, which makes the same assignments as $f$ to sequences of lengths up to $n$, and then for longer sequences simply ratifies the assignments it has made at lower levels (see the definition on p. 179).

Our set *DF* is now much more manageable than it was before we imposed our latest restriction on Type I *DF*'s. Each *DF* can now be represented as an infinite sequence of elements of [0, 1].[22] For a Type I *DF* $f$, the representation is as follows: the first member of the sequence corresponding to $f$ is the value assigned by $f$ to the empty sequence; for $i > 1$, the $i$th member of the sequence is the precision of any and all functions $N(\mathfrak{f} \circ f_{\langle a_1,\ldots,a_j \rangle})$, for $i = j + 2$.[23] For Type III *DF*'s, the representation is as follows: where $f$ is a Type I *DF* represented by the sequence $\langle f_1, f_2, f_3, \ldots \rangle$, the corresponding Type III *DF* $f_n$ (which makes the same assignments as $f$ to sequences of lengths up to $n$, and then for longer sequences, simply ratifies its own lower-level assignments) is represented by the sequence $\langle f_1, \ldots, f_{n+1}, 1, 1, 1, \ldots \rangle$. For Type II *DF*'s, **T** is represented by the sequence $\langle 1, 1, 1, \ldots \rangle$ and **F** is represented by the sequence $\langle 0, 0, 0, \ldots \rangle$.

## 5. Algebra of Degree Functions

Given the intuitive idea behind the (final, most restrictive) definition of the *DF*'s, there is one very natural choice of algebraic operations $\vee$, $\wedge$ and $'$ on *DF*. This choice derives from the natural ordering of *DF*: where $f$ and $g$ are *DF*'s represented by the sequences $\langle f_1, f_2, f_3, \ldots \rangle$ and $\langle g_1, g_2, g_3, \ldots \rangle$ respectively, we set $f \leq g$ iff $f_i \leq g_i$ for all $i$ (where the latter occurrence of '$\leq$' is the standard ordering of the reals). The thought here is as follows. For a start, we want **T** to be the maximum truth value and **F** the minimum one – and given that **T** is represented by the sequence $\langle 1, 1, 1, \ldots \rangle$ and **F** is represented by the sequence $\langle 0, 0, 0, \ldots \rangle$, this is exactly what the proposed ordering yields. Turning now to *DF*'s of Types I and III, suppose that $f$, represented as $\langle f_1, f_2, f_3, \ldots \rangle$, is the truth value of claim $P$, and $g$, represented as $\langle g_1, g_2, g_3, \ldots \rangle$, is the truth value of claim $Q$. There are two ways in which $P$ could be better than $Q$, in respect of truth: first, $P$'s degree of truth could, to a first approximation, be higher up the scale from 0 to 1; second, $P$'s degree of truth could be more localised, less diffuse. The first condition holds if $f_1$ is greater than $g_1$; the second condition holds if $f_i$ is greater than $g_i$ for *all* $i > 1$ (if $f_i > g_i$ for some $i > 1$ and $g_j > f_j$ for some $j > 1$ then $P$ and $Q$ are, overall, *incomparable* in terms of the diffuseness of their degrees of truth: one is more diffuse at one level of approximation and the other is more diffuse at another level of approximation, but overall no comparison can be drawn). Now overall, $P$ is better than $Q$, in respect of truth, only if *both* conditions hold: if $f_1$ is greater than $g_1$, but $g_j > f_j$ for some $j > 1$, then overall, $P$ and $Q$ are simply incomparable in respect of truth. Thus we arrive at the ordering of the *DF*'s stated above: where $f$ and $g$ are *DF*'s represented by the sequences $\langle f_1, f_2, f_3, \ldots \rangle$ and $\langle g_1, g_2, g_3, \ldots \rangle$ respectively, $f \leq g$ iff $f_i \leq g_i$ for all $i$.[24]

From the fact that the reals under the standard ordering form a complete lattice, it follows in a straightforward way that *DF* under the ordering just defined forms a complete lattice. Thus we are immediately given our operations $\vee$ and $\wedge$: the lattice join and meet. Where $f_i \vee g_i = \max\{f_i, g_i\}$ and $f_i \wedge g_i = \min\{f_i, g_i\}$ we define:

$$f \vee g = \langle f_1 \vee g_1, f_2 \vee g_2, f_3 \vee g_3, \ldots \rangle,$$
$$f \wedge g = \langle f_1 \wedge g_1, f_2 \wedge g_2, f_3 \wedge g_3, \ldots \rangle.$$

As for $'$, we want it to be the case that $(f')' = f$ and that if $f \leq g$, then $g' \leq f'$, and thus it is natural to make the following definition (where $f'_i = 1 - f_i$):

$$f' = \langle f'_1, f'_2, f'_3, \ldots \rangle.$$[25]

There are a number of conditions which we would like our operations on *DF* to satisfy. Three basic conditions have already been mentioned:

1. $f'(\langle\rangle) = 1 - f(\langle\rangle)$,
2. $(f \vee g)(\langle\rangle) = \max\{f(\langle\rangle), g(\langle\rangle)\}$,
3. $(f \wedge g)(\langle\rangle) = \min\{f(\langle\rangle), g(\langle\rangle)\}$.

These say "Respect the standard fuzzy operations at the level of first approximations." Apart from this, for the smooth development of blurry set theory and logic, we will want a complete bounded lattice of truth values, with an involution satisfying the De Morgan laws, and thus we require:

4. $f \vee f = f$ and $f \wedge f = f$,
5. $f \vee g = g \vee f$ and $f \wedge g = g \wedge f$,
6. $f \vee (g \vee h) = (f \vee g) \vee h$ and $f \wedge (g \wedge h) = (f \wedge g) \wedge h$,
7. $f \vee (f \wedge g) = f$ and $f \wedge (f \vee g) = f$,
8. For any set $S \subseteq DF$, there is a $g \in DF$ such that:
   (a) $\forall f \in S, f \vee g = g$
   (b) for any $h \in DF$ which possesses property (a) (i.e. $\forall f \in S, f \vee h = h$), $g \vee h = h$.
9. For any set $S \subseteq DF$, there is a $g \in DF$ such that:
   (a) $\forall f \in S, f \wedge g = g$
   (b) for any $h \in DF$ which possesses property (a) (i.e. $\forall f \in S, f \wedge h = h$), $g \wedge h = h$.
10. $f \vee \mathbf{F} = f$,
11. $f \wedge \mathbf{T} = f$,
12. $(f \vee g)' = f' \wedge g'$,
13. $(f \wedge g)' = f' \vee g'$,
14. $(f')' = f$.

Conditions 4–7 are the lattice axioms. Conditions 8 and 9 ensure that the lattice is complete: every set $S$ of *DF*'s has a supremum, which we denote $\bigvee S$, and an infimum, which we denote $\bigwedge S$. Conditions 10 and 11 ensure that $\mathbf{F}$ is an identity for $\vee$ and $\mathbf{T}$ is an identity for $\wedge$; thus the lattice is bounded. Conditions 12 and 13 are the De Morgan laws. Finally, given the other conditions, condition 14 suffices to ensure that $'$ is an involution.[26]

It is easy to confirm that our newly defined operations on *DF* satisfy all of conditions 1–14: 1–3 are immediate from the definitions of $\vee$, $\wedge$ and $'$, and 4–14 hold for the newly defined operations because they hold for the operations on the reals from which they are defined. Where $\bigvee\{f_i, g_i, h_i, \ldots\} = \sup\{f_i, g_i, h_i, \ldots\}$ and $\bigwedge\{f_i, g_i, h_i, \ldots\} = \inf\{f_i, g_i, h_i, \ldots\}$, for conditions 8 and 9 we have:

$$\bigvee\{f, g, h, \ldots\} = \left\langle \bigvee\{f_1, g_1, h_1, \ldots\}, \bigvee\{f_2, g_2, h_2, \ldots\}, \bigvee\{f_3, g_3, h_3, \ldots\}, \ldots \right\rangle,$$

$$\bigwedge\{f, g, h, \ldots\} = \left\langle \bigwedge\{f_1, g_1, h_1, \ldots\}, \bigwedge\{f_2, g_2, h_2, \ldots\}, \bigwedge\{f_3, g_3, h_3, \ldots\}, \ldots \right\rangle.$$

Given $\vee$ we may define a corresponding quotient operation: $f/g = f' \vee g$. Similarly, given $\wedge$ we may define a corresponding quotient operation: $f/g = (f \wedge g')'$. It is easy to see that the quotient operation corresponding to $\vee$ is the same as the quotient operation corresponding to $\wedge$: $f' \vee g = f' \vee g'' = (f \wedge g')'$.

We have required that if we take notice only of first approximations – that is, if we ignore everything about a *DF* except what it assigns to the empty sequence – then things should work just as they do in fuzzy logic. Equally attractive is the idea that if we take notice only of $\mathbf{T}$ and $\mathbf{F}$, then things should work just as they do in *classical* logic. We have seen that $(DF, \vee, \wedge, ', \mathbf{F}, \mathbf{T})$ is a complete bounded lattice with an involution satisfying the De Morgan laws. What we now wish to confirm is that with $\vee$, $\wedge$ and $'$ restricted to the subset $\{\mathbf{T}, \mathbf{F}\}$ of *DF*, $(\{\mathbf{T}, \mathbf{F}\}, \vee, \wedge, ')$ is a Boolean algebra, that is, a bounded distributive complemented lattice. For boundedness we want to show that for all $f$ in $\{\mathbf{T}, \mathbf{F}\}$, $\mathbf{F} \vee f = f$ and $\mathbf{T} \wedge f = f$; these facts follow immediately from conditions 4, 10 and 11. For distributivity, we wish to show that for all $f$ and $g$ in $\{\mathbf{T}, \mathbf{F}\}$, $f \vee (g \wedge h) = (f \vee g) \wedge (f \vee h)$ and $f \wedge (g \vee h) = (f \wedge g) \vee (f \wedge h)$; these facts are very easy to verify. For complementarity, we want to show that for each $f$ in $\{\mathbf{T}, \mathbf{F}\}$, there is an element $f'$ in $\{\mathbf{T}, \mathbf{F}\}$ such that $f \wedge f' = \mathbf{F}$ and $f \vee f' = \mathbf{T}$; if $\mathbf{F}' = \mathbf{T}$ and $\mathbf{T}' = \mathbf{F}$ then these facts are easy to verify – and indeed it must be the case that $\mathbf{F}' = \mathbf{T}$ and $\mathbf{T}' = \mathbf{F}$, by condition 1 and the fact that the only $DF$ which assigns 0 to $\langle\rangle$ is $\mathbf{F}$, and the only $DF$ which assigns 1 to $\langle\rangle$ is $\mathbf{T}$.

## 6. ALGEBRA OF BLURRY SETS

A blurry subset of an ordinary (or 'crisp') set $U$ is a function from $U$ to $DF$. $\mathfrak{B}U$ is the blurry power set of $U$: the set of all blurry subsets of $U$. The null or empty blurry subset $\emptyset_v$ of $U$ assigns $\mathbf{F}$ to everything in $U$; the universal blurry subset $U_v$ of $U$ assigns $\mathbf{T}$ to everything in $U$. Two blurry subsets $S_1$ and $S_2$ are identical just in case for all $x$ in $U$, $S_1(x) = S_2(x)$.[27]

Corresponding to our $DF$ algebra from the previous section is an algebra of blurry sets:

– *Complement*. For $S \in \mathfrak{B}U$, $S'$ is the complement of $S$, defined as follows: $\forall x \in U$, $S'(x) = (S(x))'$.
– *Union*. For $S_1, S_2 \in \mathfrak{B}U$, $S_1 \cup S_2$ is the union of $S_1$ and $S_2$, defined as follows: $\forall x \in U$, $(S_1 \cup S_2)(x) = S_1(x) \vee S_2(x)$. For an arbitrary family $\{S_i\}$ of sets, $(\bigcup\{S_i\})(x) = \bigvee\{S_i(x)\}$.
– *Intersection*. For $S_1, S_2 \in \mathfrak{B}U$, $S_1 \cap S_2$ is the intersection of $S_1$ and $S_2$, defined as follows: $\forall x \in U$, $(S_1 \cap S_2)(x) = S_1(x) \wedge S_2(x)$. For an arbitrary family $\{S_i\}$ of sets, $(\bigcap\{S_i\})(x) = \bigwedge\{S_i(x)\}$.
– *Containment*. For $S_1, S_2 \in \mathfrak{B}U$, $S_1 \subseteq S_2 \Leftrightarrow S_1 \cap S_2 = S_1 \Leftrightarrow S_1 \cup S_2 = S_2$. Thus for every $S \in \mathfrak{B}U$, $\emptyset_v \subseteq S$ and $S \subseteq U_v$.
– *Quotient*. For $S_1, S_2 \in \mathfrak{B}U$, $S_1/S_2$ is the quotient of $S_1$ and $S_2$, defined as follows: $\forall x \in U$, $(S_1/S_2)(x) = S_1(x)/S_2(x)$.

Given that $(DF, \vee, \wedge, ', \mathbf{F}, \mathbf{T})$ is a complete bounded lattice with an involution satisfying the De Morgan laws, it follows in a straightforward way that $(\mathfrak{B}U, \cup, \cap, ', \emptyset_v, U_v)$ is also a complete bounded lattice with an involution satisfying the De Morgan laws.

We may define an operation $^* : \mathfrak{B}U \to \mathfrak{B}U$ as follows. $\forall S \in \mathfrak{B}U$ and $\forall u \in U$:

$$S^*(u) = \begin{cases} \mathbf{F} & \text{if } S(u) \neq \mathbf{F}, \\ \mathbf{T} & \text{if } S(u) = \mathbf{F}. \end{cases}$$

$^*$ is a *pseudocomplement* because it satisfies the two conditions that, for all $S$, $S_1$ and $S_2$ in $\mathfrak{B}U$:

1. $S \cap S^* = \emptyset_v$.
2. $S_2 \subseteq S_1^* \Leftrightarrow S_1 \cap S_2 = \emptyset_v$.[28]

Let $\mathfrak{V}U^* = \{S^* \in \mathfrak{V}U : S \in \mathfrak{V}U\}$. Identifying the ordinary (crisp) subsets of $U$ with those blurry subsets which take only **T** and **F** as values, we can see that $\mathfrak{V}U^*$ consists of all the ordinary subsets of $U$. I noted at the end of Section 5 that $(\{\mathbf{T}, \mathbf{F}\}, \vee, \wedge, ')$ is a Boolean algebra. Thus with $\cup$, $\cap$ and $'$ restricted to the subset $\mathfrak{V}U^*$ of $\mathfrak{V}U$, $(\mathfrak{V}U^*, \cup, \cap, ', \emptyset_v, U_v)$ is also a Boolean algebra. We can see that the operation $'$ restricted to $\mathfrak{V}U^*$ is the same as the operation $*$ restricted to $\mathfrak{V}U^*$ (although $'$ on $\mathfrak{V}U$ is not the same as $*$ on $\mathfrak{V}U$); thus, while $*$ is a pseudo-complement on $\mathfrak{V}U$, it is a complement on $\mathfrak{V}U^*$.

## 7. Blurry Model Theory

I shall now present a blurry-set-theoretic model theory for a language. I assume that the language is a perfectly standard first-order language: the special thing about vague language is not the syntactic constructions it employs, but their meanings. The symbols of the language are as follows:

– the propositional connectives $\neg$, $\vee$, $\wedge$, and $\rightarrow$
– the quantifiers $\exists$ and $\forall$
– the punctuation marks ( , ) and ,
– infinitely many variables $x_1, x_2, \ldots$
– infinitely many individual constants $a_i$ and predicate letters $A^n_k$ (for $i, n, k \geq 1$; superscripts represent number of arguments; subscripts are index numbers).

Terms and well-formed formulae (wfs) are defined in the usual way.

A blurry interpretation $\mathfrak{M} = (M, I)$ of this language consists in the following:

– A nonempty set $M$ (the domain)
– An interpretation function $I$ which assigns:
  - to each individual constant $a_k$, an object $(a_k)^{\mathfrak{M}} \in M$
  - to each $n$-adic predicate letter $A^n_k$, a function $(A^n_k)^{\mathfrak{M}} : M^n \rightarrow DF$ (i.e. a blurry subset of $M^n$)

Terminology:

– $[\mathcal{A}]_{\mathfrak{M}}$ is the truth value of $\mathcal{A}$ on interpretation $\mathfrak{M}$; this value is a member of $DF$
– $\mathcal{A}_y a$ is the sentence obtained by writing $a$ in place of all free occurrences of $y$ in $\mathcal{A}$, $a$ being some constant that does not occur in $\mathcal{A}$
– $\mathfrak{M}^a_o$ is the interpretation which is just like $\mathfrak{M}$ except that in it the constant $a$ is assigned the denotation $o$.

The truth values of closed wfs on an interpretation are defined as follows:

1. $[A^n_k(a_i, \ldots, a_j)]_{\mathfrak{M}} = (A^n_k)^{\mathfrak{M}}((a_i)^{\mathfrak{M}}, \ldots, (a_j)^{\mathfrak{M}})$
2. $[\neg\mathcal{A}]_{\mathfrak{M}} = ([\mathcal{A}]_{\mathfrak{M}})'$
3. $[\mathcal{A} \vee \mathcal{B}]_{\mathfrak{M}} = [\mathcal{A}]_{\mathfrak{M}} \vee [\mathcal{B}]_{\mathfrak{M}}$
4. $[\mathcal{A} \wedge \mathcal{B}]_{\mathfrak{M}} = [\mathcal{A}]_{\mathfrak{M}} \wedge [\mathcal{B}]_{\mathfrak{M}}$
5. $[\mathcal{A} \rightarrow \mathcal{B}]_{\mathfrak{M}} = [\mathcal{A}]_{\mathfrak{M}}/[\mathcal{B}]_{\mathfrak{M}}$
6. $[\exists y\mathcal{A}]_{\mathfrak{M}} = \bigvee(\{[\mathcal{A}_y a]_{\mathfrak{M}^a_o} : o \in M\})$
7. $[\forall y\mathcal{A}]_{\mathfrak{M}} = \bigwedge(\{[\mathcal{A}_y a]_{\mathfrak{M}^a_o} : o \in M\})$.

### 7.1. *Identity*

I have not yet mentioned the relation of *identity*. The topic of vague identity – and more generally, of vague objects – is one about which there is much to be said, but a proper discussion of this topic is beyond the scope of this paper: my focus here is on vague predication and vague properties. There are in fact interesting relationships between the views developed in this paper and the issues surrounding vague objects, but also, there is no *essential* connection between them: we could combine the view of vague properties and predication developed here with a classical view of identity. Suppose that we expand our language to include the binary predicate symbol '$=$'; we could then add the following clause to the definition of truth for closed wfs:

8. $[a_i = a_j]_{\mathfrak{M}} = \begin{cases} \mathbf{T} & \text{if } (a_i)^{\mathfrak{M}} = (a_j)^{\mathfrak{M}}, \\ \mathbf{F} & \text{otherwise.} \end{cases}$

In the absence of a proper discussion of the topic of vague objects, this may serve as a default position on identity.

## 8. CONSEQUENCE

I shall use the notation $\langle\mathcal{A}\rangle_{\mathfrak{M}}$ as an abbreviation of $[\mathcal{A}]_{\mathfrak{M}}(\langle\rangle)$, i.e. the value assigned to the empty sequence by the truth value (i.e. *DF*) of $\mathcal{A}$ on interpretation $\mathfrak{M}$. I shall say that $\mathcal{B}$ is a blurry consequence of a set $\Gamma$ of wfs ($\Gamma \models_v \mathcal{B}$) just in case there is no interpretation $\mathfrak{M}$ such that $\langle\mathcal{A}\rangle_{\mathfrak{M}} > 0.5$, for every $\mathcal{A}$ in $\Gamma$, and $\langle\mathcal{B}\rangle_{\mathfrak{M}} < 0.5$. Correspondingly, I shall say that $\mathcal{B}$ is a blurry tautology ($\models_v \mathcal{B}$) just in case there is no interpretation $\mathfrak{M}$ such that $\langle\mathcal{B}\rangle_{\mathfrak{M}} < 0.5$. There are two points to note about this definition of consequence. First, in deciding whether $\mathcal{B}$ is a blurry consequence of $\mathcal{A}$, we look at all possible interpretations of $\mathcal{A}$ and $\mathcal{B}$, but on each interpretation, all that we are concerned with is *the value assigned to the empty sequence* by $\mathcal{A}$'s and $\mathcal{B}$'s *DF*'s. In other words, consequence is decided at the level of first approximations. Second, the definition cannot be recast in terms of *preservation of designated values* (in this case, of values assigned to the empty sequence): for it to be the case that $\mathcal{A} \models_v \mathcal{B}$, it is required that on any interpretation on which the value assigned by $\mathcal{A}$'s *DF* to the empty sequence is *strictly greater* than 0.5, the value assigned by $\mathcal{B}$'s *DF* to the empty sequence is *greater than or equal to* 0.5. I shall discuss these points below; first, I shall show that the blurry consequence relation on the language specified above is identical to the classical consequence relation ($\models$) on that language.

LEMMA 1. $\models \mathcal{A} \Rightarrow \models_v \mathcal{A}$.

*Proof.* For any blurry interpretation $\mathfrak{M} = (M, I)$ of our language there is a corresponding classical interpretation $\mathfrak{M}_c = (M_c, I_c)$, specified as follows. $M_c = M$, and for $I_c$:

- $(a_k)^{\mathfrak{M}_c} = (a_k)^{\mathfrak{M}}$.
- Where $I$ assigns to an $n$-adic predicate letter $A^n_k$ a function $(A^n_k)^{\mathfrak{M}}$ : $M^n \to DF$, $I_c$ assigns to $A^n_k$ a function $(A^n_k)^{\mathfrak{M}_c} : (M_c)^n \to \{0, 1\}$, as follows. For each $n$-tuple $x$ in $M^n$, let $f_x$ be the *DF* assigned to $x$ by $(A^n_k)^{\mathfrak{M}}$. If $f_x(\langle\rangle) \geq 0.5$, then $(A^n_k)^{\mathfrak{M}_c}(x) = 1$; if $f_x(\langle\rangle) < 0.5$ then $(A^n_k)^{\mathfrak{M}_c}(x) = 0$.[29]

The truth values of closed formulae on this classical interpretation are determined by the usual classical valuation rules.

We can prove by induction on complexity of sentences that for any blurry interpretation $\mathfrak{M}$ and any sentence $\mathcal{A}$, if $\langle\mathcal{A}\rangle_{\mathfrak{M}} < 0.5$ then $\mathcal{A}$ is false on the corresponding classical interpretation $\mathfrak{M}_c$, and if $\langle\mathcal{A}\rangle_{\mathfrak{M}} > 0.5$ then $\mathcal{A}$ is true on the corresponding classical interpretation $\mathfrak{M}_c$.[30]

*Base*: atomic sentences. If $\langle A^n_k(a_i, \ldots, a_j)\rangle_{\mathfrak{M}} < 0.5$, then $(A^n_k)^{\mathfrak{M}}((a_i)^{\mathfrak{M}}, \ldots, (a_j)^{\mathfrak{M}})(\langle\rangle) < 0.5$, so $(A^n_k)^{\mathfrak{M}_c}((a_i)^{\mathfrak{M}_c}, \ldots, (a_j)^{\mathfrak{M}_c}) = 0$, so $A^n_k(a_i, \ldots, a_j)$ is false on $\mathfrak{M}_c$. If $\langle A^n_k(a_i, \ldots, a_j)\rangle_{\mathfrak{M}} > 0.5$, then $(A^n_k)^{\mathfrak{M}}((a_i)^{\mathfrak{M}}, \ldots, (a_j)^{\mathfrak{M}})(\langle\rangle) > 0.5$, so $(A^n_k)^{\mathfrak{M}_c}((a_i)^{\mathfrak{M}_c}, \ldots, (a_j)^{\mathfrak{M}_c}) = 1$, so $A^n_k(a_i, \ldots, a_j)$ is true on $\mathfrak{M}_c$.

*Induction*: one case for each of clauses 2–7 of the valuation scheme. In fact we need not check the clauses for $\wedge$, $\rightarrow$ and $\forall$, because these could just as well have been introduced by definition, given $\neg$, $\vee$ and $\exists$.[31]

- If $\langle\neg\mathcal{A}\rangle_{\mathfrak{M}} < 0.5$, then $\langle\mathcal{A}\rangle_{\mathfrak{M}} > 0.5$, so by the induction hypothesis, $\mathcal{A}$ is true on $\mathfrak{M}_c$, hence $\neg\mathcal{A}$ is false on $\mathfrak{M}_c$. If $\langle\neg\mathcal{A}\rangle_{\mathfrak{M}} > 0.5$, then $\langle\mathcal{A}\rangle_{\mathfrak{M}} < 0.5$, so by the induction hypothesis, $\mathcal{A}$ is false on $\mathfrak{M}_c$, hence $\neg\mathcal{A}$ is true on $\mathfrak{M}_c$.

- If $\langle \mathcal{A} \vee \mathcal{B} \rangle_{\mathfrak{M}} < 0.5$, then $\langle \mathcal{A} \rangle_{\mathfrak{M}} < 0.5$ and $\langle \mathcal{B} \rangle_{\mathfrak{M}} < 0.5$, so by the induction hypothesis, $\mathcal{A}$ is false on $\mathfrak{M}_c$ and $\mathcal{B}$ is false on $\mathfrak{M}_c$, so $\mathcal{A} \vee \mathcal{B}$ is false on $\mathfrak{M}_c$. If $\langle \mathcal{A} \vee \mathcal{B} \rangle_{\mathfrak{M}} > 0.5$, then $\langle \mathcal{A} \rangle_{\mathfrak{M}} > 0.5$ or $\langle \mathcal{B} \rangle_{\mathfrak{M}} > 0.5$, so by the induction hypothesis, $\mathcal{A}$ is true on $\mathfrak{M}_c$ or $\mathcal{B}$ is true on $\mathfrak{M}_c$, so $\mathcal{A} \vee \mathcal{B}$ is true on $\mathfrak{M}_c$.
- If $\langle \exists y \mathcal{A} \rangle_{\mathfrak{M}} < 0.5$, then $\langle \mathcal{A}_y a \rangle_{\mathfrak{M}^a_o} < 0.5$ for every $o$ in $M$, so by the induction hypothesis, $\mathcal{A}_y a$ is false on every interpretation $(\mathfrak{M}^a_o)_c$; but for every $o$, $(\mathfrak{M}^a_o)_c = (\mathfrak{M}_c)^a_o$, so $\exists y \mathcal{A}$ is false on $\mathfrak{M}_c$. If $\langle \exists y \mathcal{A} \rangle_{\mathfrak{M}} > 0.5$, then $\langle \mathcal{A}_y a \rangle_{\mathfrak{M}^a_o} > 0.5$ for some $o$ in $M$, so by the induction hypothesis, $\mathcal{A}_y a$ is true on some interpretation $(\mathfrak{M}^a_o)_c = (\mathfrak{M}_c)^a_o$, so $\exists y \mathcal{A}$ is true on $\mathfrak{M}_c$.

Thus for any sentence $\mathcal{A}$, if there is a blurry interpretation $\mathfrak{M}$ such that $\langle \mathcal{A} \rangle_{\mathfrak{M}} < 0.5$, then there is a classical interpretation on which $\mathcal{A}$ is false. Contraposing, if $\mathcal{A}$ is a classical tautology (i.e. there is no classical interpretation on which $\mathcal{A}$ is false) then $\mathcal{A}$ is a blurry tautology (i.e. there is no blurry interpretation $\mathfrak{M}$ such that $\langle \mathcal{A} \rangle_{\mathfrak{M}} < 0.5$). □

LEMMA 2. $\Gamma, \mathcal{A} \models_v \mathcal{B} \Leftrightarrow \Gamma \models_v \mathcal{A} \rightarrow \mathcal{B}$.

*Proof.* ($\Leftarrow$) Suppose $\Gamma \models_v \mathcal{A} \rightarrow \mathcal{B}$. Consider an arbitrary interpretation $\mathfrak{M}$ such that $\langle \gamma \rangle_{\mathfrak{M}} > 0.5$ for all $\gamma$ in $\Gamma$, and $\langle \mathcal{A} \rangle_{\mathfrak{M}} > 0.5$; then $\langle \neg\mathcal{A} \rangle_{\mathfrak{M}} < 0.5$. By supposition $\langle \mathcal{A} \rightarrow \mathcal{B} \rangle_{\mathfrak{M}} = \langle \neg\mathcal{A} \vee \mathcal{B} \rangle_{\mathfrak{M}} \geq 0.5$, hence $\langle \mathcal{B} \rangle_{\mathfrak{M}} \geq 0.5$.

($\Rightarrow$) Suppose $\Gamma, \mathcal{A} \models_v \mathcal{B}$. Consider an arbitrary interpretation $\mathfrak{M}$ such that $\langle \gamma \rangle_{\mathfrak{M}} > 0.5$ for all $\gamma$ in $\Gamma$. If $\langle \mathcal{A} \rangle_{\mathfrak{M}} > 0.5$ then by supposition $\langle \mathcal{B} \rangle_{\mathfrak{M}} \geq 0.5$, hence $\langle \mathcal{A} \rightarrow \mathcal{B} \rangle_{\mathfrak{M}} = \langle \neg\mathcal{A} \vee \mathcal{B} \rangle_{\mathfrak{M}} \geq 0.5$. If $\langle \mathcal{A} \rangle_{\mathfrak{M}} \leq 0.5$ then $\langle \neg\mathcal{A} \rangle_{\mathfrak{M}} \geq 0.5$, hence $\langle \mathcal{A} \rightarrow \mathcal{B} \rangle_{\mathfrak{M}} = \langle \neg\mathcal{A} \vee \mathcal{B} \rangle_{\mathfrak{M}} \geq 0.5$. □

THEOREM 1. $\Gamma \models \mathcal{A} \Leftrightarrow \Gamma \models_v \mathcal{A}$.

*Proof.* ($\Rightarrow$) Suppose $\Gamma \models \mathcal{A}$; then there is some finite $\gamma = \{\mathcal{A}_1, \ldots, \mathcal{A}_n\} \subseteq \Gamma$ such that $\gamma \models \mathcal{A}$; hence $\mathcal{A}_1 \wedge \cdots \wedge \mathcal{A}_n \models \mathcal{A}$; hence $\models (\mathcal{A}_1 \wedge \cdots \wedge \mathcal{A}_n) \rightarrow \mathcal{A}$. So by Lemma 1, $\models_v (\mathcal{A}_1 \wedge \cdots \wedge \mathcal{A}_n) \rightarrow \mathcal{A}$, hence by Lemma 2, $\mathcal{A}_1 \wedge \cdots \wedge \mathcal{A}_n \models_v \mathcal{A}$. If there were a blurry interpretation $\mathfrak{M}$ such that $\langle \mathcal{B} \rangle_{\mathfrak{M}} > 0.5$ for all $\mathcal{B}$ in $\Gamma$ and $\langle \mathcal{A} \rangle_{\mathfrak{M}} < 0.5$, then in particular $\langle \mathcal{A}_i \rangle_{\mathfrak{M}} > 0.5$ for all $\mathcal{A}_i$ in $\gamma$, hence $\langle \mathcal{A}_1 \wedge \cdots \wedge \mathcal{A}_n \rangle_{\mathfrak{M}} > 0.5$, contradicting the fact that $\mathcal{A}_1 \wedge \cdots \wedge \mathcal{A}_n \models_v \mathcal{A}$; so there is no such interpretation, i.e. $\Gamma \models_v \mathcal{A}$.

($\Leftarrow$) As noted at the end of Section 5, $(\{\mathbf{T}, \mathbf{F}\}, \vee, \wedge, ')$ is a Boolean algebra; hence any classical interpretation is a special case of a blurry interpretation (thinking of 0 or The False as **F** and 1 or The True as **T**). A classical interpretation on which every $\gamma$ in $\Gamma$ is true and $\mathcal{A}$ is false is a blurry interpretation on which every $\gamma$ in $\Gamma$ has truth value **T** and $\mathcal{A}$

has truth value **F**. Thus if there is no blurry interpretation $\mathfrak{M}$ such that $\langle\gamma\rangle_{\mathfrak{M}} > 0.5$ for every $\gamma$ in $\Gamma$ and $\langle\mathcal{A}\rangle_{\mathfrak{M}} < 0.5$, then *a fortiori* there is no classical interpretation on which every $\gamma$ in $\Gamma$ is true and $\mathcal{A}$ is false. □

The blurry consequence relation on the language defined above is, then, identical to the classical consequence relation on that language. We thus have a non-classical semantics giving rise to classical logic. In this respect we are in the same position as the supervaluationist,[32] and this position is widely felt to be superior to the positions of those who advocate non-classical logics of vagueness. It must be noted, however, that the result about consequence cuts only so much ice. For a start, while our first-order language is expressively complete with respect to classical truth functions, it is *not* expressively complete with respect to *DF*-valued truth functions. This would only be a significant issue, however, if further truth-functions were needed in order to express our ordinary reasoning with vague concepts: and I do not believe that this is the case. Second, we will later introduce some truth predicates, and these *will* be needed in order to express some of our ordinary claims about vagueness. The supervaluationist faces a similar issue, however: in order to express some of our ordinary claims about vagueness, she needs *definitely* and *indefinitely* operators (or truth predicates).

Given the result about consequence, the question of *proof theory* for blurry logic is straightforward. Any proof theory for our first-order language that is sound and complete with respect to classical set-theoretic models is sound and complete with respect to blurry set-theoretic models: for example, any standard classical proof theory, whether in the axiomatic, natural deduction, sequent calculus or tableaux style.

I shall now discuss the two points raised at the beginning of this section. The first point is that consequence is decided at the level of first approximations. One might wonder: why introduce degree functions in all their complexity, if everything but the assignment to the empty sequence is ignored when it comes to the question of consequence? Well, the point of introducing degree functions in all their complexity is to accommodate the phenomenon of higher-order vagueness, failure to accommodate which was the downfall of the fuzzy account. (I shall explain in Section 11 exactly how my account deals with higher-order vagueness.) When it comes to logical consequence, however, we certainly do *not* want the higher levels of degree functions playing a central role. This becomes clear when we consider the notion of *soundness*. A valid argument is one such that, whenever the premises have a certain special property *P* which premises may have, the conclusion has a certain special property *C* which conclusions may have; a *sound* argument is one which is valid, and whose premises do indeed have this special property $P$ (and hence whose conclusion must have property $C$). $P$ and $C$ are usually the same property; for example in classical logic they are the property *truth*. In general, however, we need not assume that they are the same – an issue which I shall discuss shortly. Now soundness is supposed to be a *useful* notion: it should *not* be defined in such a way that it is, in principle, impossible for us to determine whether any piece of ordinary reasoning is sound. But if the definition of consequence referred to degree functions in all their glorious complexity, then this is precisely what would happen: we would in general be unable to determine of a piece of ordinary reasoning involving vague concepts whether it was sound. In order to do so, we would need to determine whether the conclusion was a consequence of the premises – whether if the premises all had $P$, the conclusion would have $C$ – *and* we would need to determine whether the premises did indeed have $P$. But in the envisioned circumstances, $P$ and $C$ involve degree functions in all their complexity, whereas in general for any vague sentence $S$, all but the lowest levels of $S$'s degree function are unknown to ordinary speakers. Given a vague sentence $S$, we can in general hazard a first approximation to its degree of truth (that is, we can hazard a fuzzy degree), and this will in general be reasonably close to the value which its actual truth value (a degree function) assigns to the empty sequence. Thus, first approximations are accessible to ordinary speakers, and a definition of $P$ and $C$, and hence of consequence and of soundness, may reasonably make reference to them. However in general, beyond first approximations the degree functions of the sentences we use are unknown to us: the detail at higher levels is there precisely in order to allow room for the thought that what we do have access to is *not* the full and final story of the degree of truth of the sentences we use. Hence in defining validity and soundness we should restrict ourselves to properties $P$ and $C$ that are decided at the level of first approximations.[33]

What about the second point mentioned at the beginning of this section: that the definition of consequence allows that in a valid argument, the premises might be more true (to a first approximation) than the conclusion? Why adopt this definition, rather than – as is standard practice in many-valued logics – a definition in terms of preservation of designated values? Why have $P$ and $C$ being different properties? Obviously part of the answer here is: "Because my definition works." My definition yields a classical consequence relation, and this is important. Epistemicists have lorded their classicism over their opponents, and supervaluationists have claimed their greater adherence to classical principles as an advantage over their fuzzy rivals. An important constraint on a definition of validity is that it counts intuitively valid forms of reasoning as valid – and the classically valid inference forms are all (*pace* relevant logicians, and other non-classical logicians whose motivation is primarily proof-theoretic) *prima facie* paradigms of valid forms of reasoning, even in contexts involving vagueness. But of course, the 'because it works' answer is not enough: if it were, we could give whatever semantics we pleased for a language, and then simply say, "$S$ is a consequence of $\Gamma$ just in case it is a classical consequence." What would be missing here is a meaningful relationship between the semantics and the definition of consequence.

In the present case, however, there *is* a meaningful relationship between the blurry set-theoretic semantics and the definition of blurry consequence. Let us begin with the notion of a tautology. One semantic property that classical tautologies possess is the property of being true (i.e. having the truth value The True) on every interpretation; another is the property of having the maximum truth value on every interpretation; a third is the property that a sentence $S$ has just in case, on every interpretation, $S$ is at least as true as $\neg S$ (in other words, on every interpretation, a tautology is at least as true as its negation). How are we to generalise the notion of a tautology to the case of blurry semantics, where the latter two properties – which are equivalent in the classical context – come apart? Well, the general rule in this type of situation is that we should pick whichever of the generalisations is most useful. Consider for example the notion of one set of objects *having fewer elements* than another set. In finite sets, if $U$ is obtained from $V$ by removing some members of $V$, then $U$ has fewer elements than $V$, while if $U$ and $V$ can be mapped onto each other, then they contain the same number of members. Now consider infinite sets, and in particular the set of positive integers and the set of even positive integers. The latter can be obtained from the former by removing some members, but on the other hand, each set can be mapped onto the other. So do they have the same number of members, or are there fewer even positive integers than positive integers? Two conflicting answers suggest themselves, depending upon which of the facts about the finite case we take as our point of departure. In this situation it was quite acceptable for (Cantor, 1915) to choose the most useful generalisation – the one that led to the better overall theory of transfinite numbers – and that is just what he did.[34] Returning to the case of vagueness, we face a similar type of choice concerning the definition of a blurry tautology, and we should choose the most useful generalisation.[35] We are under no obligation to say that a tautology has the value **T** on every interpretation: we could say this, but then, the property of having the value **T** on every interpretation is not very interesting, because so few sentences have it. We get a better overall theory if we generalise the other idea, that a tautology is always at least as true as its negation. Confining ourselves to first approximations, this becomes: to a first approximation, a tautology is always at least as true as its negation; or equivalently, a tautology is at least 0.5 true (to a first approximation) on every interpretation – which is the definition I chose. On this way of looking at things, if you utter a tautology $S$, you may not be uttering something that is as true as true can be (i.e. that has the maximum truth value), but you will certainly be making a *safe* choice: you could never say something more true by uttering the negation of $S$ instead.

With this definition in place, we know what property $C$ must be in the definition of consequence: it must be the tautology property[36] – the property of being at least 0.5 true, to a first approximation – for we wish to retain the idea that an argument from no premises, with a tautology as conclusion, is valid, and the idea that the conclusion of a valid argument with no premises is a tautology. What about property $P$? In the classical framework, we may say that $P$ and $C$ are both the property of being *true*; or we may say that $P$ is the property of being *true enough to form the basis of a sound piece of reasoning*, while $C$ is the property of being *true enough to assert safely*. Given that the only truth values available are The True and The False, these amount to the same thing. They may not amount to the same thing in the blurry framework, however. Suppose we adopt the option $P = C$ (where $C$ is the property of being at least 0.5 true, to a first approximation). It will then turn out that many paradigms of valid inference are invalid, for example *modus ponens*: for we can have $\langle \mathcal{A} \rangle_{\mathfrak{M}} \geq 0.5$ and $\langle \mathcal{A} \to \mathcal{B} \rangle_{\mathfrak{M}} \geq 0.5$ and yet $\langle \mathcal{B} \rangle_{\mathfrak{M}} = 0$. *Modus ponens* seems like a paradigm of valid inference *even in vague contexts* – so we should consider the option of making the requirements on possessing the property *true enough to form the basis of a sound piece of reasoning* more stringent than the requirements on possessing the property *true enough to assert safely*. The idea would be that a sentence needs to meet more stringent standards of truth if it is to be used as the basis for further argument than if it is merely to be asserted – just as building codes place more stringent standards of load-bearing capacity on foundations than on superstructures.

Now at this point I still have not given any reason for making the requirements on possessing the property *true enough to form the basis of a sound piece of reasoning* more stringent than the requirements on possessing the property *true enough to assert safely*, other than that this may yield a classical consequence relation. But in fact, some idea of this sort is very natural in the context of vagueness. Intuitively, the more steps you take down a Sorites series, the more shaky your conclusion (e.g., that the man before you is bald) becomes. Given that man 1 is bald, it is safe to say that man 2 is bald – but maybe not safe enough for you to then go on and assert that because man 2 is bald, man 3 is bald also. When we first encounter the Sorites paradox, we feel that from the fact that man 1 is bald, and that successive men in the series differ by just a hair, we may conclude that man 2 is bald – and man 3, and perhaps man 4 – but that the further we progress along the series, the shakier the conclusion becomes. Some sentences are, so to speak, true enough to get a terminating pass – they may be asserted, but not built on – while other sentences have a higher grade of truth, and may be used as the basis of further reasoning. So in vague contexts, the distinction between sentences which are true enough to form the basis of a sound piece of reasoning, and sentences which are merely true enough to assert safely, is a natural one. Given this, and the fact that Sorites reasoning seems perfectly valid, it is natural to say that a valid argument is one in which, if the premises are true enough to be used as premises, then the conclusion is true enough to be asserted safely. Now for the reasons given above, we want this property 'true enough to form the basis of a sound piece of reasoning' to make reference only to first approximations. Should we then say that a sentence is true enough to form the basis of a sound piece of reasoning if it is 1 true, to a first approximation? This seems too demanding: we can reason soundly in vague contexts, even though none of our statements are 1 true, to a first approximation. So how true is true enough? Well at *this* point, the fact that if we say 'true enough' is 'strictly greater than 0.5 true' then we get the classical consequence relation, is good reason to say it – and hence, finally, we arrive at the definition of blurry consequence given at the beginning of this section.

The fact that a blurrily valid inference always takes us from sentences which are true enough to form the basis of a sound piece of reasoning to sentences which are true enough to assert safely – rather than to sentences which are also true enough to form the basis of a sound piece of reasoning – does not mean that the relation of blurry consequence is not transitive. In fact it *is* transitive: it must be, because it is the same relation as the classical consequence relation, and that is transitive![37]

### 8.1. *Conditionals*

Related to the question of consequence is the question of the conditional. The semantics for the conditional given above does not generalise the standard fuzzy semantics for the conditional, according to which $[\mathcal{A} \rightarrow \mathcal{B}] = 1 + [\mathcal{B}] - \max\{[\mathcal{A}], [\mathcal{B}]\}$. The main reason for my choice is that I want the blurry consequence relation to be the same as the classical consequence relation (provided this can be achieved with a reasonable definition of blurry consequence – but as discussed in the previous section, it can be) and I want to retain the usual connection between consequence and the conditional: $\mathcal{B}$ is a consequence of $\mathcal{A}$ just in case $\mathcal{A} \rightarrow \mathcal{B}$ is a tautology (i.e. is a consequence of the empty set of sentences). Given my definitions of consequence and the conditional, we do indeed have this connection, as Lemma 2 shows. If we were to adopt instead a generalisation of the standard fuzzy conditional, then we might be able to redefine consequence in such a way as to regain the connection – but then our consequence relation could not (as far as I can see) be classical.

Nevertheless, it might be thought: doesn't the standard fuzzy account provide a better formal rendition of the English 'if ... then ...' than the account presented above, according to which $\mathcal{A} \rightarrow \mathcal{B}$, $\neg\mathcal{A} \vee \mathcal{B}$ and $\neg(\mathcal{A} \wedge \neg\mathcal{B})$ always have the same truth value? For example, consider Bob, a borderline case of 'bald', and Bill, who has one less hair than Bob. Let us suppose 'Bob is bald' is 0.5 true (i.e. to a first approximation) and 'Bill is bald' is 0.51 true. Then 'If Bob is bald, then Bill is bald' is 0.51 true, according to my semantics, whereas on the analogue of the standard fuzzy semantics, it would be 1 true – and isn't the latter the more intuitive assignment? In fact this is not clear. In saying 'If Bob is bald, then Bill is bald' one might mean that if one were to stipulate a sharp boundary for 'bald', and it enclosed Bob, then it must enclose Bill also – i.e. one is saying that if Bob *counts as* bald, then Bill *counts as* bald. This is certainly something I want to accept – and I can easily accept it, for this claim about boundary stipulation is *not* properly (semi-)formalised as 'Bob is bald $\rightarrow$ Bill is bald'. On the other hand, in saying 'If Bob is bald, then Bill is bald' one might be saying simply that if Bob is bald, then Bill is bald (just as one might say 'If the United States wins the World Cup, then the profile of soccer in this country will increase dramatically'). In this case it does *not* seem that the sentence should be definitely true. Suppose I am unwrapping my Christmas presents; I get to a longish object and say 'If this is a spade I will use it to dig a vegetable garden.' It turns out to be a two-piece fishing rod, and looking at it, I say again 'If this is a spade I will use it to dig a vegetable garden.' This is simply a very odd thing to say – for we can all clearly see that it is *not* a spade. Now suppose you ask me whether Bob and Bill – neither of whom I have seen – are bald, and you tell me that Bill has one less hair than Bob. I say 'Well then, if Bob is bald, then Bill is bald.' Now Bob and Bill are brought in, and I see that Bob is a borderline case for 'bald'. Now if I say again 'If Bob is bald, then Bill is bald' (and I mean just what I say) then far from being clearly true, my statement is simply odd: not as odd as the spade statement – because it is not *clearly false* that Bob is bald – but quite odd nevertheless. What I say makes perfect sense if it means 'If Bob *counts as* bald, then Bill *counts as* bald'; and I could also say quite truly, before seeing the men, 'If Bob is *clearly* bald, then Bill is bald' – but then upon seeing them, it would be odd to repeat this; but as for plain old 'If Bob is bald, then Bill is bald', the less true its antecedent, the odder a thing it is to say (provided, as in the case just presented, all parties to the discourse know the truth-status of the antecedent).[38]

I do not think, then, that the fuzzy conditional is clearly a better rendition of 'if . . . then . . .' than the conditional defined above – or at least, the example considered does not show that it is. But if you disagree, then you should consider the following example. Suppose that 'Ben is tall' is 0.51 true. Then, using the fuzzy conditional, both 'If Bob is bald, then Bill is bald' and 'If Bob is bald, then Ben is tall' are true to degree 1. Presumably, however, someone who believes that the former is true to a high degree will not also believe that the latter is true to a high degree.[39] Thus, from the point of view of someone who believes that the former is true to a high degree, the fuzzy conditional has taken us from the frying pan, but only as far as the fire.

We could go on considering examples and pumping intuitions, but this would not get us very far. The conditional is one of the most versatile and puzzling of English constructions, and the question of its formalisation is a very complex one, to which many works have been devoted. But not only would a full treatment of the conditional occupy far more space than I have available: there is in any case no reason to think that a logic of *vagueness* should shed any special light on the question of conditionals.[40] Thus it is best if we admit that no truth definition for the conditional within a degree-theoretic account of vagueness is going to accommodate all our intuitions about conditionals, or be the final answer to the question of the proper formalisation of 'if . . . then . . .'. The right thing to do is to define, as one's basic conditional, a connective that behaves reasonably well, and leave the matter there, with no claim to have cast any new light on the broader questions about conditionals. This is the approach that I have taken: my conditional has many nice properties, and while some persons' intuitions may count against it, it is certainly not an unacceptable rendition of 'if . . . then . . .'.

There are two points worth noting before we leave this issue. First, we do not have to confine ourselves to one conditional. We could define as many others as we please (but if we were to do so, the question of proof theory would need to be addressed anew). Second, there is an important point of connection between the conditional defined above and the standard fuzzy conditional. In the standard fuzzy account, the tautology property – the property that a tautology has on *every* interpretation – is having the value 1, and in the case of the standard fuzzy conditional, if $[\mathcal{A}] \leq [\mathcal{B}]$, then $[\mathcal{A} \rightarrow \mathcal{B}] = 1$, that is, the conditional has the tautology property. In the account presented in this paper, the tautology property is having a value of at least 0.5 (to a first approximation), and in the case of the conditional defined above, if $\langle\mathcal{A}\rangle \leq \langle\mathcal{B}\rangle$, then $\langle\mathcal{A} \rightarrow \mathcal{B}\rangle \geq 0.5$,[41] that is, the conditional has the tautology property.

Note that the foregoing account of consequence and the conditional could be applied to the fuzzy theory: if we abandon the standard fuzzy conditional in favour of a conditional like the one defined above, and abandon the definition of fuzzy consequence in terms of preservation of designated values, in favour of a definition analogous to that given above, then we get a fuzzy consequence relation on our first-order language that is identical to the classical consequence relation on that language.

## 9. Speaking of Truth

In motivating my proposal for a formal model of the intuitive notion of *degree* of possession of a property, I appealed to the intuition that when I say that Bob is bald to degree 0.7, this is just an approximation, and you might well say that my statement is true to degree 0.8 – and someone might say that your statement is true to degree 0.4, and so on. So we might well ask how such locutions as ‘Bill’s statement is true to degree 0.3’ can be handled within (an extension of) the formal framework presented above.

Consider the classical framework for a moment. Suppose that you say “Bob is bald”, and I say “That’s true”. There are two things I might be doing. I might be saying, in effect, “Ditto”, that is, reasserting what you asserted. In this case my statement should have the same truth value as yours. On the other hand I might be making an assertion about your statement, namely the assertion that your statement has the property of being true. In this case, if your statement is indeed true, then mine should be true, and if your statement is false, then my statement should be false. But that is just to say that my statement should have the same truth value as yours, so from the semantic point of view, there is no difference between the truth predicate I use when I reassert what you say, and the truth predicate I use when I say, of your statement, that it has the property of being true.

In non-classical frameworks, on the other hand, there may well be a semantic difference between truth predicates as used to perform these different conversational tasks. In the framework presented in this paper we may distinguish three sorts of truth predicate. First, there is the ‘ditto’ predicate, $T$. If you say "Bob is bald" and I say, in the 'ditto' sense, "That's true", then my statement should have the same truth value (i.e. *DF*) as yours. Second, there are the approximate truth predicates, $T_x$: one for each $x \in [0, 1]$. If you say "Bob is bald" and I make the claim, concerning your statement, that it is true to degree 0.3 – i.e. $T_{0.3}$ – then if your statement has the *DF* $f$, my statement should have the *DF* $f/0.3$, which assigns to the empty sequence what $f$ assigns to $\langle 0.3 \rangle$, assigns to 0.2 what $f$ assigns to $\langle 0.3, 0.2 \rangle$, and so on. In general, if $s$ is a sentence with *DF* $f$, then the sentence $T_x s$ has *DF* $f/x$, defined as follows:

$$f/x(\langle s_1, \ldots, s_n \rangle) = f(\langle x, s_1, \ldots, s_n \rangle).$$

Third, there are the definite truth predicates, $T_f$: one for each $f \in DF$. If you say "Bob is bald" and I make the claim, concerning your statement, that it has the *DF* $f$, then if your statement does indeed have $f$ as its truth value, then my statement has *DF* **T**, while if your statement has a *DF* other than $f$, then my statement has *DF* **F**.

The approximate truth predicates $T_x$ are the ones of most interest in modelling our ordinary use of vague predicates, but it is possible to consistently introduce all the predicates just mentioned into the formal framework presented above, and I shall now do so. The method employed is based upon the construction in (Kripke, 1975).

I begin by introducing a new *DF*: $*$. Intuitively a sentence with $*$ as its *DF* suffers from a truth value gap. We might (but need not) think of $*$ as the totally partial function from $[0, 1]^*$ to $[0, 1]$: the function that assigns no value to any sequence. (Recall that all the other *DF*'s are total functions from $[0, 1]^*$ to $[0, 1]$.) On notation: *DF* does not include $*$; let $DFG = DF \cup \{*\}$. I shall call $*$ a 'value', but reserve the name 'truth value' for members of *DF*. Likewise, I shall reserve the name 'degree' for members of *DF*, $*$ being thought of as a degree gap.

I now need to extend the valuation scheme presented earlier to handle cases in which some sentences have no truth value (i.e. have value $*$). The extension is by analogy with Kleene's weak three valued logic, which is to say that $*$ trumps.[42] Originally I presented clauses for $\neg$, $\vee$, $\wedge$, $\rightarrow$, $\exists$ and $\forall$. These clauses are not all necessary: given $\neg$ and $\vee$ and $\exists$, $\wedge$ and $\rightarrow$ and $\forall$ can be introduced by definition. So I now consider only the first three cases. The new valuation scheme is as follows. The definition of an interpretation is the same as before, except that this time, the interpretation function $I$ assigns to each $n$-adic predicate letter $A^n_k$ a function from $M^n$ to *DFG* (not *DF*), and $[\mathcal{A}]_{\mathfrak{M}}$ – the value of $\mathcal{A}$ on interpretation $\mathfrak{M}$ – is a member of *DFG* (not *DF*). The values of closed formulae on an interpretation are defined as follows:

(1) $[A^n_k(a_i, \dots, a_j)]_{\mathfrak{M}} = (A^n_k)^{\mathfrak{M}}((a_i)^{\mathfrak{M}}, \dots, (a_j)^{\mathfrak{M}}),$

(2) $[\neg\mathcal{A}]_{\mathfrak{M}} = \begin{cases} * & \text{if } [\mathcal{A}]_{\mathfrak{M}} = *, \\ ([\mathcal{A}]_{\mathfrak{M}})' & \text{otherwise,} \end{cases}$

(3) $[\mathcal{A} \vee \mathcal{B}]_{\mathfrak{M}} = \begin{cases} * & \text{if } [\mathcal{A}]_{\mathfrak{M}} = * \text{ or } [\mathcal{B}]_{\mathfrak{M}} = *, \\ [\mathcal{A}]_{\mathfrak{M}} \vee [\mathcal{B}]_{\mathfrak{M}} & \text{otherwise,} \end{cases}$

(4) $[\exists y\mathcal{A}]_{\mathfrak{M}} = \begin{cases} * \quad \text{if } [\mathcal{A}_y a]_{\mathfrak{M}^a_o} = * \text{ for any } o \in M, \\ \bigvee\{[\mathcal{A}_y a]_{\mathfrak{M}^a_o} : o \in M\} \quad \text{otherwise.} \end{cases}$

Say that an interpretation $\mathfrak{M}'$ *extends* an interpretation $\mathfrak{M}$ (in symbols $\mathfrak{M} \leq \mathfrak{M}'$ or $\mathfrak{M}' \geq \mathfrak{M}$) iff (i) they have the same domain $M$, and assign the same denotations to all names, and (ii) for every $n$-adic predicate $A^n$ and every $x \in M^n$ and any $f \in DF$ (note: not $DFG$), if $(A^n)^{\mathfrak{M}}(x) = f$ then $(A^n)^{\mathfrak{M}'}(x) = f$. Clause (ii) says that if, according to interpretation $\mathfrak{M}$, object $x$ has property $A$ to some degree $f$ (where $f$ is not $*$, this being thought of as a degree gap), then according to interpretation $\mathfrak{M}'$, object $x$ has property $A$ to that same degree $f$. Thus $\mathfrak{M}'$ might classify objects that $\mathfrak{M}$ left unclassified, but if $\mathfrak{M}$ classifies an object – i.e. associates it with some predicate to some degree – then $\mathfrak{M}'$ classifies it in the same way.

LEMMA 3. *The valuation scheme presented above is monotone; that is, for any sentence $\phi$ and any $f \in DF$ (note: not DFG):*

$$\mathfrak{M} \leq \mathfrak{M}' \Rightarrow ([\phi]_{\mathfrak{M}} = f \Rightarrow [\phi]_{\mathfrak{M}'} = f).$$

That is, monotonicity of the valuation scheme says that if $\mathfrak{M}'$ extends $\mathfrak{M}$, then for any sentence, if $\mathfrak{M}$ assigns it a truth value $f$ (i.e. a value other than $*$) then $\mathfrak{M}'$ assigns it the same truth value.

*Proof.* Monotonicity is proved by induction on complexity of sentences:

*Base*: atomic sentences $A^n(a_i, \dots, a_j)$. Suppose $[A^n(a_i, \dots, a_j)]_{\mathfrak{M}} = f$. By clause 1 of the valuation scheme, $(A^n)^{\mathfrak{M}}((a_i)^{\mathfrak{M}}, \dots, (a_j)^{\mathfrak{M}}) = f$. But supposing $\mathfrak{M} \leq \mathfrak{M}'$, $(a_k)^{\mathfrak{M}} = (a_k)^{\mathfrak{M}'}$ by clause (i) of the definition of extension, and so by clause (ii) of the definition of extension, $(A^n)^{\mathfrak{M}'}((a_i)^{\mathfrak{M}'}, \dots, (a_j)^{\mathfrak{M}'}) = f$. Thus by clause 1 of the valuation scheme, $[A^n(a_i, \dots, a_j)]_{\mathfrak{M}'} = f$.

*Induction*: non-atomic sentences. There is a case for each of clauses 2–4 of the valuation scheme.

(i) $\neg\mathcal{A}$. Suppose $[\neg\mathcal{A}]_{\mathfrak{M}} = f$. By clause 2 of the valuation scheme, $[\mathcal{A}]_{\mathfrak{M}} = g$, where $f = g'$. So by the induction hypothesis, $[\mathcal{A}]_{\mathfrak{M}'} = g$, whence by clause 2 of the valuation scheme, $[\neg\mathcal{A}]_{\mathfrak{M}'} = g' = f$.

(ii) $\mathcal{A} \vee \mathcal{B}$. Suppose $[\mathcal{A} \vee \mathcal{B}]_{\mathfrak{M}} = f$. By clause 3 of the valuation scheme, $[\mathcal{A}]_{\mathfrak{M}} = f_1$ and $[\mathcal{B}]_{\mathfrak{M}} = f_2$, where $f = f_1 \vee f_2$. So by the induction hypothesis, $[\mathcal{A}]_{\mathfrak{M}'} = f_1$ and $[\mathcal{B}]_{\mathfrak{M}'} = f_2$, whence by clause 3 of the valuation scheme, $[\mathcal{A} \vee \mathcal{B}]_{\mathfrak{M}'} = f_1 \vee f_2 = f$.[43]

(iii) $\exists y\mathcal{A}$. Suppose $[\exists y\mathcal{A}]_{\mathfrak{M}} = f$. By clause 4 of the valuation scheme, for every $o \in M$, $[\mathcal{A}_y a]_{\mathfrak{M}^a_o} \neq *$, and $\bigvee\{[\mathcal{A}_y a]_{\mathfrak{M}^a_o} : o \in M\} = f$. Now $\mathfrak{M}' \geq \mathfrak{M}$, so for every $o \in M$, $\mathfrak{M}'^a_o \geq \mathfrak{M}^a_o$. Hence by the induction hypothesis, for every $o \in M$, $[\mathcal{A}_y a]_{\mathfrak{M}'^a_o} = [\mathcal{A}_y a]_{\mathfrak{M}^a_o}$. So $\bigvee\{[\mathcal{A}_y a]_{\mathfrak{M}'^a_o} : o \in M\} = f$. So $[\exists y\mathcal{A}]_{\mathfrak{M}'} = f$. □

Before introducing the truth predicates mentioned earlier, we begin with an interpreted language $L$ – interpreted in accordance with the model theory presented above. Let the domain of the interpretation be $M$. $M$ will remain fixed throughout the following construction. We suppose that all predicates are totally defined, in the sense that none of them is assigned an extension which assigns $*$ to any object in $M$ (or in general in $M^n$, for an $n$-ary predicate). We now extend $L$ to $\mathcal{L}$ by adding new monadic predicates $T$, $T_x$ and $T_f$: one $T_x$ for each $x \in [0, 1]$ and one $T_f$ for each $f \in DF$. Initially these predicates are uninterpreted, but the ultimate point of the exercise is to establish that they may be regarded as, respectively, a 'ditto' truth predicate ($T$), approximate truth predicates (the $T_x$'s), and definite truth predicates (the $T_f$'s).

For now, however, some more preliminaries. We define an order relation $\leq$ on $DFG$ as follows: $\forall x, y \in DFG, x \leq y$ iff (i) $x = y$ or (ii) $x = *$ and $y \in DF$. That is, $*$ is less than everything, while each member of $DF$ is less than nothing except itself. Let $\langle DFG, \leq\rangle$ be $DFG$ together with the relation just defined. Obviously $\langle DFG, \leq\rangle$ is a partial order.

Where $\langle D, \leq\rangle$ is a partial order, call $X \subseteq D$ *consistent* if for every $x, y \in X$ there is a $z \in D$ such that $x \leq z$ and $y \leq z$. Then $\langle D, \leq\rangle$ is a *coherent complete partial order* or *ccpo* if every consistent $X \subseteq D$ has a supremum in $D$.[44] Obviously $\langle DFG, \leq\rangle$ is a ccpo.

$DFG^M$ is the set of all functions from the domain $M$ to the set $DFG$, that is, the set of all blurry subsets of $M$ (including partial subsets: ones which assign $*$ to some objects). For $U, V \in DFG^M$, set $U \leq V$ just in case $U(x) \leq V(x)$ for all $x \in M$ (where the most recent occurrence of '$\leq$' denotes the relation on $DFG$ defined above).[45] Because $\langle DFG, \leq\rangle$ is a ccpo, so is $\langle DFG^M, \leq\rangle$.[46]

Let $I = [0, 1] \cup DF \cup \{t\}$, where $t$ is some arbitrary object that is not in $[0, 1]$ or $DF$. Consider $(DFG^M)^I$, the set of all functions from $I$ to $DFG^M$.

(Think of $I$ as a set of names, and each member of $(DFG^M)^I$ as a *named set of blurry subsets* of $M$. For $F \in (DFG^M)^I$, we have one subset $F(t)$ whose name is $t$, a subset $F(x)$ named $x$ for each $x \in [0, 1]$, and a subset $F(f)$ named $f$ for each $f \in DF$.) For $F, G \in (DFG^M)^I$, set $F \leq G$ just in case $F(i) \leq G(i)$ for all $i \in I$. Then, just as before, $\langle (DFG^M)^I, \leq \rangle$ is a ccpo.

For $F \in (DFG^M)^I$, let $\mathcal{L}(F)$ be the interpretation of $\mathcal{L}$ in which $T$ is assigned $F(t)$ as its extension, $T_x$ is assigned $F(x)$ as its extension (for each $x \in [0, 1]$), and $T_f$ is assigned $F(f)$ as its extension (for each $f \in DF$).

We now define a function $'$ on $(DFG^M)^I$ as follows. (We write $'(F)$ as $F'$. Note that $F'(i)$ is the subset of $M$ named $i$, under the naming system $F'$. We can specify $F'$ by specifying $F'(i)$ for each $i \in I$, i.e. for $i = t$, $i = x \in [0, 1]$ and $i = f \in DF$. In turn we can specify each $F'(i)$ by specifying the member of $DFG$ that it assigns to each $y \in M$. Recall that the notation '$f/x$' is defined on p. 203.) For all $y \in M$:

| | | | |
|---|---|---|---|
| $F'(t)(y)$ | $=$ | $\mathbf{F}$ | if $y$ is not a sentence |
| | $=$ | $*$ | if $y$ is a sentence, and $y$ is assigned $*$ in $\mathcal{L}(F)$ |
| | $=$ | $f$ | if $y$ is a sentence, and $y$ is assigned $f \in DF$ in $\mathcal{L}(F)$ |
| $F'(x)(y)$ | $=$ | $\mathbf{F}$ | if $y$ is not a sentence |
| | $=$ | $*$ | if $y$ is a sentence, and $y$ is assigned $*$ in $\mathcal{L}(F)$ |
| | $=$ | $f/x$ | if $y$ is a sentence, and $y$ is assigned $f \in DF$ in $\mathcal{L}(F)$ |
| $F'(f)(y)$ | $=$ | $\mathbf{F}$ | if $y$ is not a sentence, or if $y$ is a sentence, and $y$ is assigned $g \neq f \in DF$ in $\mathcal{L}(F)$ |
| | $=$ | $*$ | if $y$ is a sentence, and $y$ is assigned $*$ in $\mathcal{L}(F)$ |
| | $=$ | $\mathbf{T}$ | if $y$ is a sentence, and $y$ is assigned $f \in DF$ in $\mathcal{L}(F)$. |

LEMMA 4. *If $F \leq G$ then $\mathcal{L}(F) \leq \mathcal{L}(G)$ (i.e. $\mathcal{L}(G)$ extends $\mathcal{L}(F)$).*

*Proof.* The only difference between $\mathcal{L}(F)$ and $\mathcal{L}(G)$ is in the extensions they assign to $T$, the $T_x$'s and the $T_f$'s. We thus need only show that for any one $T_i$ of these predicates (thinking of $T$ as $T_t$, for the sake of convenience of presentation) and any $y \in M$ and any $f \in DF$, if $(T_i)^{\mathcal{L}(F)}(y) = f$ then $(T_i)^{\mathcal{L}(G)}(y) = f$. Now $(T_i)^{\mathcal{L}(F)} = F(i)$ and $(T_i)^{\mathcal{L}(G)} = G(i)$, so if $F \leq G$, the desired result follows (recall that in the ordering of $DFG$ specified earlier, the only thing that $f$ bears $\leq$ to is $f$). □

THEOREM 2. *$'$ is a monotone function on $(DFG^M)^I$, i.e. $F \leq G \Rightarrow (F' \leq G')$.*

*Proof.* $F' \leq G'$ iff $F'(i) \leq G'(i)$ for all $i \in I$, and $F'(i) \leq G'(i)$ iff $F'(i)(y) \leq G'(i)(y)$ for all $y \in M$. Pick an arbitrary $i$ and $y$. Case (i): If

$F'(i)(y) = *$ then $F'(i)(y) \leq G'(i)(y)$, whatever $G'(i)(y)$ is. Case (ii): $F'(i)(y) = f \in DF$. Case (ii-a): $y$ is not a sentence. Then $F'(i)(y) = \mathbf{F} = G'(i)(y)$. Case (ii-b): $y$ is a sentence. We need to consider the three cases $i = t$, $i = x \in [0, 1]$ and $i = f \in DF$:

(1) $i = t$. $F'(i)(y) = f$, so $y$ is assigned $f$ in $\mathcal{L}(F)$. Then by Lemmas 3 and 4, $y$ is assigned $f$ in $\mathcal{L}(G)$. Hence $G'(i)(y) = f$.
(2) $i = x \in [0, 1]$. $F'(i)(y) = f$, so $y$ is assigned $g$ in $\mathcal{L}(F)$, where $f = g/x$. Then by Lemmas 3 and 4, $y$ is assigned $g$ in $\mathcal{L}(G)$. Hence $G'(i)(y) = g/x = f$.
(3) $i = f \in DF$. There are two cases to consider:
    (a) $F'(i)(y) = \mathbf{F}$, so $y$ is assigned $g \neq f$ in $\mathcal{L}(F)$. Then by Lemmas 3 and 4, $y$ is assigned $g$ in $\mathcal{L}(G)$. Hence $G'(i)(y) = \mathbf{F}$.
    (b) $F'(i)(y) = \mathbf{T}$, so $y$ is assigned $f$ in $\mathcal{L}(F)$. Then by Lemmas 3 and 4, $y$ is assigned $f$ in $\mathcal{L}(G)$. Hence $G'(i)(y) = \mathbf{T}$. □

We now define $\circledast \in DFG^M$ to be the totally partial blurry subset of $M$: the one which assigns $*$ to every object in $M$. Let $F^{\circledast} \in (DFG^M)^I$ be the function which assigns $\circledast$ to every $i \in I$.

THEOREM 3. *$F^{\circledast}$ is a sound point of the function $'$, that is:*

$$F^{\circledast} \leq F^{\circledast\prime}.$$

*Proof.* $F^{\circledast} \leq F^{\circledast\prime}$ iff $F^{\circledast}(i) \leq F^{\circledast\prime}(i)$ for all $i \in I$, and $F^{\circledast}(i) \leq F^{\circledast\prime}(i)$ iff $F^{\circledast}(i)(y) \leq F^{\circledast\prime}(i)(y)$ for all $y \in M$. But $F^{\circledast}(i)(y) = *$ for all $i \in I$ and for all $y \in M$, so whatever $F^{\circledast\prime}(i)(y)$ is, $F^{\circledast}(i)(y) \leq F^{\circledast\prime}(i)(y)$. □

Now we can link in with a general theorem which says that any monotone function on a ccpo with a sound point has a fixed point. We need two lemmas. In all cases we are concerned with a monotone function $f$ on a ccpo $\langle X, \leq\rangle$, where $X$ is a set.[47]

LEMMA 5. *If $x \in X$ is a sound point of $f$, then $f(x)$ is a sound point of $f$.*

*Proof.* $x$ is a sound point of $f$, i.e. $x \leq f(x)$, so by monotonicity $f(x) \leq f(f(x))$, i.e. $f(x)$ is a sound point of $f$. □

LEMMA 6. *If $Y \subseteq X$ is a consistent set of sound points of $f$ then the supremum $\bigvee Y$ of $Y$ is a sound point of $f$.*

*Proof.* We want to show that $\bigvee Y \leq f(\bigvee Y)$. This will follow if $f(\bigvee Y)$ is an upper bound of $Y$, because $\bigvee Y$ is $Y$'s least upper bound. So we wish to show that for any $x \in Y$, $x \leq f(\bigvee Y)$. Well, $x \leq \bigvee Y$ by definition of $\bigvee Y$, so $f(x) \leq f(\bigvee Y)$ by monotonicity of $f$, and $x \leq f(x)$ because $x$ is a sound point, so $x \leq f(\bigvee Y)$ by transitivity of $\leq$. □

THEOREM 4. *If $x \in X$ is a sound point of $f$, then $f$ has a fixed point $x' \in X$ (i.e. a point $x'$ such that $f(x') = x'$) with $x \leq x'$.*

*Proof.* Let $x_0 = x$, and for each successor ordinal $\alpha + 1$ define $x_{\alpha+1} = f(x_\alpha)$, and for each limit ordinal $\lambda$ define $x_\lambda = \bigvee_{\beta<\lambda} x_\beta$. We show by induction on $\alpha$ that each $x_\alpha$ does in fact exist, and is a sound point of $f$. $x_0$ exists and is sound by supposition. $x_{\alpha+1}$ exists by the supposition that $f$ is defined on $X$, and by Lemma 5, $x_{\alpha+1}$ is sound if $x_\alpha$ is. Suppose that $\lambda$ is a limit ordinal and that $x_\beta$ is defined and sound for all $\beta < \lambda$. Then $x_\beta \leq x_{\beta+1}$ for all such $\beta$ (because each $x_\beta$ sound), hence $x_\beta \leq x_\delta$ for all $\beta < \delta < \lambda$. Thus for any $x_\beta$ and $x_\delta$ in $\{x_\beta : \beta < \lambda\}$, either $x_\beta \leq x_\delta$ or $x_\delta \leq x_\beta$. Thus $\{x_\beta : \beta < \lambda\}$ is consistent, and so by definition of a ccpo, $x_\lambda = \bigvee_{\beta<\lambda} x_\beta$ exists, and by Lemma 6 is sound. So $x_\alpha$ exists and is sound for all $\alpha$, and if $\alpha < \beta$ then $x_\alpha \leq x_\beta$. Now either $x_\alpha < x_{\alpha+1}$ for all $\alpha$, or there is some $\alpha$ such that $x_\alpha = x_{\alpha+1}$. In the former case, $\{x_\alpha : \alpha \in \text{ON}\}$ is in one-to-one correspondence with ON (the class of all ordinals), which is impossible because $X$ is a set. So there is some $\alpha$ such that $x_\alpha = x_{\alpha+1} = f(x_\alpha)$, i.e. $x_\alpha$ is a fixed point of $f$; and $x = x_0 \leq x_\alpha$ because $0 \leq \alpha$. □

We now have everything we need to get an interpretation of our new predicates $T$, the $T_x$'s and the $T_f$'s. We specify a hierarchy of interpretations $\mathcal{L}_\alpha$ of $\mathcal{L}$, for ordinal $\alpha$ (we know from the proof of Theorem 4 that the hierarchy is well defined):

1. $\mathcal{L}_0 = \mathcal{L}(F_0) = \mathcal{L}(F^{\circledast})$
2. $\mathcal{L}_{\alpha+1} = \mathcal{L}(F_{\alpha+1}) = \mathcal{L}(F_\alpha{}')$
3. For limit $\lambda$, $\mathcal{L}_\lambda = \mathcal{L}(F_\lambda) = \mathcal{L}(\bigvee_{\beta<\lambda} F_\beta)$.

Theorems 2, 3 and 4 tell us that there is an $\alpha$ such that $F_\alpha = F_{\alpha+1}$. The interpretation $\mathcal{L}_\alpha = \mathcal{L}_{\alpha+1}$ is then one in which the sentence $Ts$ has the same value as $s$ for every sentence $s$ (and has value **F** where $s$ is not a sentence), the sentence $T_x s$ has value $f/x$ where $s$ is a sentence with truth value $f$ (and has value **F** where $s$ is not a sentence, and value $*$ where $s$ is a sentence with value $*$), and the sentence $T_f s$ has value **T** where $s$ is a sentence with truth value $f$ and value **F** where $s$ is a sentence with truth value $g \neq f$ (and has value **F** where $s$ is not a sentence, and value $*$ where $s$ is a sentence with value $*$), which is exactly what we wanted.[48]

## 10. Denying Bivalence

Williamson has argued that bivalence cannot coherently be denied: any denial of bivalence implies a contradiction.[49] As an advocate of a view which denies bivalence, I need to consider his argument. Where $\mathbf{a}$ is a name of the sentence $\mathcal{A}$, $\bar{\mathbf{a}}$ is a name of the sentence $\neg\mathcal{A}$ and $\mathbb{T}$ is a truth predicate, the argument is as follows (Williamson, 1992, pp. 265–266):

| | | |
|---|---|---|
| 1. | $\neg(\mathbb{T}\mathbf{a} \vee \mathbb{T}\bar{\mathbf{a}})$ | [the denial of bivalence], |
| 2. | $\mathbb{T}\mathbf{a} \leftrightarrow \mathcal{A}$ and $\mathbb{T}\bar{\mathbf{a}} \leftrightarrow \neg\mathcal{A}$ | [two instances of Tarski's T-schema], |
| 3. | $\neg(\mathcal{A} \vee \neg\mathcal{A})$ | [from 1 and 2], |
| 4. | $\neg\mathcal{A} \wedge \neg\neg\mathcal{A}$ | [from 3]. |

As (Williamson, 1992, pp. 266–267, n. 4) notes, in order for this argument to carry weight, step 2 must be validated, and step 4 must indeed be absurd.

In my framework, there are three truth predicates which we could consider in place of $\mathbb{T}$ in the above argument: $T$, $T_1$ and $T_{\mathbf{T}}$. I shall consider these in reverse order.

In the case of $T_{\mathbf{T}}$, step 2 is not validated. $T_{\mathbf{T}}\mathbf{a} \leftrightarrow \mathcal{A}$ has the same *DF* as $(\neg T_{\mathbf{T}}\mathbf{a} \vee \mathcal{A}) \wedge (\neg\mathcal{A} \vee T_{\mathbf{T}}\mathbf{a})$, which is not a tautology: if $\mathcal{A}$ is 0.9 true (i.e. to a first approximation), then $T_{\mathbf{T}}\mathbf{a}$ has *DF* **F**; so the right conjunct is 0.1 true and the left conjunct is 1 true, so the whole conjunction is 0.1 true.

In the case of $T_1$ also, step 2 is not validated. $T_1\mathbf{a} \leftrightarrow \mathcal{A}$ has the same *DF* as $(\neg T_1\mathbf{a} \vee \mathcal{A}) \wedge (\neg\mathcal{A} \vee T_1\mathbf{a})$, which is not a tautology: if $\mathcal{A}$ is 0.7 true (i.e. to a first approximation), and to a second approximation, it is 0.01 true that $\mathcal{A}$ is 1 true, then $T_1\mathbf{a}$ is 0.01 true; so the left conjunct is 0.99 true and the right conjunct is 0.3 true; so the whole conjunction is 0.3 true.

In the case of $T$, step 2 *is* validated: $T$ *does* satisfy the T-schema. $T\mathbf{a} \leftrightarrow \mathcal{A}$ has the same *DF* as $(\neg T\mathbf{a} \vee \mathcal{A}) \wedge (\neg\mathcal{A} \vee T\mathbf{a})$, which *is* a tautology: for it to be less than 0.5 true (i.e. to a first approximation), one of its conjuncts must be less than 0.5 true, hence both disjuncts of this conjunct must be less than 0.5 true – but this is impossible, because on every (fixed point) interpretation, $\mathcal{A}$ and $T\mathbf{a}$ have the same *DF*, so one disjunct is the negation of something which has the same value as the other disjunct. The problem with $T$, however, is that the version of step 1 involving $T$ is *not* a denial of bivalence, and is not something to which the proponent of many-valued semantics is committed. Williamson's dialectic is as follows: "You, the proponent of many-valued semantics, are committed to asserting a sentence which denies bivalence; but this sentence implies a contradiction, so you are committed to asserting a contradiction."[50] But suppose that Bob is a borderline case of baldness, and so we wish to deny that 'Bob is bald' is true or false. If '**a**' names this sentence, and '$\bar{\mathbf{a}}$' names its negation, then we should go about this denial by asserting $\neg(T_{\mathbf{T}}\mathbf{a} \vee T_{\mathbf{T}}\bar{\mathbf{a}})$ or $\neg(T_1\mathbf{a} \vee T_1\bar{\mathbf{a}})$, *not* by asserting $\neg(T\mathbf{a} \vee T\bar{\mathbf{a}})$. The latter is not something to which the advocate of many-valued semantics is committed. This advocate might well wish to say *in a tentative, hedging way* that it is not the case that Bob is bald nor that he is not bald, but if so, he would be just as happy to say, in the same tentative, hedging way, that Bob is bald but also not bald. These are just the sorts of things non-philosophers do say about borderline cases.[51] So in this third case Williamson's argument goes through: if we are committed to step 1, we are committed to step 4. The problem for Williamson is that no-one need be committed to step 1. We might well be *sort of* committed to step 1, but the fact that we are then also *sort of* committed to step 4 is no revelation, and no problem.

It seems then that Williamson's error was to ignore the following point. In a many-valued semantic framework, there will be one truth predicate for each truth value, and a 'ditto' truth predicate. On the one hand, the former do not obey the Tarskian T-schema, while the latter does: in general there is no reason why '$S$ has truth value $x$' should have the same truth value as $S$, but the 'ditto' predicate is precisely the predicate '$T$' such that '$S$ is $T$' and $S$ always have the same truth value. On the other hand, the denial of bivalence that characterises the many-valued system will be formalised in terms of one of the *former* predicates: the proponent of the many-valued system is committed to the idea that in general, for a sentence $S$ and a truth value $x$, it does not have to be the case that either $S$ has $x$ or $S$'s negation has $x$. So the proponent of the many-valued semantics is immune from Williamson's reductio, which requires that *the truth predicate used to deny bivalence* satisfies the T-schema.

## 11. Vagueness and Higher-Order Vagueness

Having now presented my view, it is time to check that it does what we want it to do. At the outset, I mentioned two problems for the fuzzy view: the problem of the linear ordering of fuzzy truth values, and the higher-order vagueness problem. Obviously my view is immune to the linear ordering worry: the natural ordering of *DF* is *not* linear. Thus two issues remain to be discussed: we need to verify that the account of vagueness presented in this paper has the resources to accommodate both vagueness (conceived in terms of Closeness), and the higher-order vagueness that the fuzzy account cannot accommodate.

11.1. *Vagueness*

According to the Closeness characterisation, a predicate '$F$' is vague just in case it satisfies the condition that if $a$ and $b$ are very similar in $F$-relevant respects, then '$Fa$' and '$Fb$' are very similar in respect of truth. Now in a sorites series for the predicate $F$, adjacent members of the series *are* very similar in $F$-relevant respects (for example, in a series for 'tall' they differ by a millimetre in height; in a series for 'bald' they differ by a hair; etc.). For vagueness to be accommodated and the sorites paradox avoided, we need it to be the case that sentences of the form '$Fx$', where '$x$' refers to a member of the sorites series for $F$, can start out being true simpliciter (where '$x$' refers to the object at one end of the series), end up being false simpliciter (where '$x$' refers to the object at the other end of the series), *and* be such that '$Fx$' and '$Fx'$' are always very similar in respect of truth, when '$x$' and '$x'$' refer to adjacent members of the series. Now the fuzzy view, for example, clearly has the resources to achieve this result: we can start at one end of the series with '$Fx$' being 1 true, and move to the other end of the series at which '$Fx$' is 0 true, in such a way that the difference between the truth values of '$Fx$' and '$Fx'$' is always very small.

What about the blurry account? Does it have the resources to accommodate vagueness as characterised in terms of Closeness? Clearly it does. We start at one end of the series with '$Fx$' having **T** as its truth value; and where '$x$' refers to the second object in the series, '$Fx$' has as its truth value any *DF* $f$, represented as $\langle f_1, f_2, f_3, \ldots\rangle$, where $f_1$ is a number very close, but not equal, to 1 (say 0.999), and each other $f_i$ is a number very close, or equal, to 1: intuitively, any such *DF* is very similar to **T** in respect of truth. At the other end of the series, '$Fx$' will have **F** as its truth value; and where '$x$' refers to the second last object in the series, '$Fx$' has as its truth value any *DF* $f$, represented as $\langle f_1, f_2, f_3, \ldots\rangle$, where $f_1$ is a number very close, but not equal, to 0 (say 0.001), and each other $f_i$ is a number very close, or equal, to 1: intuitively, any such *DF* is very similar to **F** in respect of truth. In between, we can progress in such a way that where $f$, represented as $\langle f_1, f_2, f_3, \ldots\rangle$, is the truth value assigned to '$Fx$', and $f'$, represented as $\langle f'_1, f'_2, f'_3, \ldots\rangle$, is the truth value assigned to '$Fx'$', it is the case that for each $i$, the difference between $f_i$ and $f'_i$ is extremely small – and hence intuitively, $f$ and $f'$ are very similar truth values. Thus we may move between **T** and **F** in a series of steps in which we only ever pass from a *DF* to a *DF* that is very close to it in respect of truth – and thus we do have the resources to accommodate predicates which satisfy Closeness.

### 11.2. *Finite Higher-Order Vagueness*

The problem of higher-order vagueness for the fuzzy account was that no sentence of the form '$S$ is $x$ true' (where $x$ is an element of [0, 1]) can have an intermediate degree of truth. In presenting normal degree functions, we found it natural to extend the 'higher-order vagueness' terminology in the following way: $S$ is first-order vague if $S$ has an intermediate degree of truth; $S$ is second-order vague if some sentence of the form '$S$ is $x$ true' (where $x$ is an element of [0, 1]) has an intermediate degree of truth; $S$ is third-order vague if some sentence of the form ' '$S$ is $x$ true' is $y$ true' has an intermediate degree of truth (where $x$ and $y$ are elements of [0, 1]); and so on. The fuzzy account cannot accommodate sentence vagueness above the first order. It is clear from the discussion of the approximate truth predicates $T_x$ in Section 9, however, that the account presented in this paper can accommodate sentence vagueness of *all* finite orders. If $S$ has $f$ as its truth value, where $f$ is a Type I *DF*, then $S$ is vague at every finite order; if $S$ has $f_n$ as its truth value, where $f_n$ is a Type III *DF*, then $S$ is vague up to and including the $(n + 1)$th order, but no higher (recall that $n$ is a non-negative integer); and if $S$ has a Type II *DF* (**T** or **F**) as its truth value, then $S$ is not vague at all. Thus we see that the higher-order vagueness problem for the fuzzy account does not pose a problem for the account presented in this paper. Furthermore – as mentioned at the outset – the problem is avoided without positing a hierarchy of vague metalanguages. There is a hierarchical structure in my account, but it is *inside* each truth value: each truth value is, as it were, the limit of a sequence of approximations of itself; it contains within itself an infinite hierarchy of such approximations.

### 11.3. *Transfinite Higher-Order Vagueness*

It is natural to extend our terminology of sentence vagueness in the following way: a sentence $S$ is $\omega$-order vague if some sentence of the form '$S$ is $f$ true' (where $f$ is a *DF*) has an intermediate degree of truth. It is clear from the discussion in Section 9 that the blurry account does *not* accommodate $\omega$-order sentence vagueness: sentences of the form $T_f a$ always have **T** or **F** as their truth value.[52] Is it a problem for my view that it does not accommodate transfinite higher order vagueness? No, it is not. I shall now consider some reasons why someone might think it *is* a problem, and show what is wrong with these reasons.

First, someone might think:

> It was unacceptable that the fuzzy theorist gave us just one option regarding the degree of truth of 'Bob is bald': we don't want to be told that it is definitely true that Bob is 0.6 bald, and definitely false that he is

0.60001 bald. But on your account, 'Bob is bald' has a certain *DF* as its truth value, and this *DF* assigns one particular value to each sequence. Consider, for example, the empty sequence: according to your account, the *DF* assigns (say) 0.6 to it, and not 0.60001 or any other number, and thus, to a first approximation, 'Bob is bald' is 0.6 true, and not 0.6001 true. But surely if Bob is approximately 0.6 bald, he is *also* approximately 0.60001 bald. Thus your view is just as arbitrary and unacceptable as the fuzzy view.

There are two misunderstandings of my view wrapped up in this objection. First, I fully agree that if Bob is approximately 0.6 bald, then he is also approximately 0.60001 bald. This fact is reflected in my account in the fact that if you say 'Bob is 0.6 bald' and I say 'Bob is 0.60001 bald', our statements can be similar in respect of truth. It is not that there is just one correct first approximation to the degree of truth of 'Bob is bald', this being the value assigned to the empty sequence by the *DF* of 'Bob is bald'. Rather, there are as many first approximations as there are numbers in [0, 1], and some of them are better than others. The value assigned to the empty sequence by the *DF* of 'Bob is bald' should be thought of not as *the* first approximation to the degree of truth of 'Bob is bald', but as the *canonical* first approximation. What distinguishes this approximation from the others is simply that this one lies at the centre of mass of the density function given at the next level of approximation. Of course, if I were to say not that 'Bob is bald' is (approximately) 0.6 true, but that the *canonical* approximation to the degree of truth of 'Bob is bald' is 0.6, then what I said would be either definitely true or definitely false: this is because what I said would be equivalent to saying that the *DF* of 'Bob is bald' assigns 0.6 to the empty sequence, which is a statement about *DF*'s. The issue under discussion is precisely whether such statements should indeed always be definitely true or false; my aim at present is to clear up the misunderstandings embodied in the first attempt to argue that they should not be. The first misunderstanding, then, was to think that on my view, there is one correct first approximation to Bob's degree of baldness: there is not; there are many approximations, some better and some worse, but none uniquely correct.

With this misunderstanding cleared up, the objection presented above amounts to the following:

It seems arbitrary that the fuzzy account assigns a particular value in [0, 1] to each statement. But on your account, each sequence in [0, 1]* is assigned a particular value in [0, 1], and this seems just as arbitrary! Why should it be the case that the *DF* of 'Bob is bald' assigns 0.8 to some sequence, rather than 0.80001? Why should this or that density function have its centre of mass right *here*, and why should it have precisely *this* shape?; and so on . . .

The intuitions behind this objection are intuitions which I share: they are in fact intuitions about *finite* higher order vagueness, and they *are* accommodated by my account. Certainly it seems wrong to say (as the fuzzy account says) that Bob is bald to precisely *this* fuzzy degree: this is an intuition about the existence of second-order vagueness, and is accommodated in my account by the fact that a *DF* assigns values not just to the empty sequence, but to sequences of length 1. Of course, it seems arbitrary in turn to say that it is precisely $x$ true that Bob is $y$ bald: this is an intuition about the existence of third-order vagueness, and is accommodated in my account by the fact that a *DF* assigns values not just to the empty sequence and to sequences of length 1, but to sequences of length 2. The curve given at the second level of approximation is blurred at the *third* level of approximation: there is no need to go to transfinite levels in order to blur it. And so on: the intuition that there is something rough or arbitrary or approximate or incomplete about the assignment to any sequence $\langle a_1 \ldots, a_n \rangle$ is accommodated *within* my view in the assignments of values to sequences $\langle a_1 \ldots, a_n, x \rangle$: we do not need to go back again, after all the assignments are fixed, and do something *more* to accommodate this intuition. Positing transfinite levels would be overkill: the tasks the objector thinks they are needed for have all already been performed *within* the existing hierarchy of approximations; there is nothing more that needs to be blurred out.

A second sort of objector accepts my response to the first objection, but asserts that just as the fuzzy theory was wrong to associate a particular number in [0, 1] with each sentence, so I am wrong to associate a particular degree function with each sentence: the objection is that I face the very problem that the fuzzy theory faces, only at a higher level – I have not avoided the problem of higher-order vagueness, I have merely postponed it.

There is a tendency amongst philosophers to reflexively reapply a form of argument that works in one context, in any context that resembles the first context. Thus, in the present case, the reflex reaction would be as follows: "The fuzzy theorist assigned a unique degree of truth to each sentence, and it got her into trouble; you assign a unique degree of truth to each sentence, so you get into the same – or similar – trouble." This sort of reflex argument carries no weight, by itself: we need to actually check that the new argument against my view has the same intuitive force as the original argument against the fuzzy view. The intuition embodied in the higher-order vagueness objection to the fuzzy account is widespread: the fuzzy account fails to accommodate a genuine, robust intuition about vagueness.[53] Is there also a genuine, robust intuition to the effect that some or all vague predicates are such that some sentences in which they figure are vague not only at all finite orders, but at some transfinite order? I cannot see that there is.

We need to distinguish two scenarios. In the first scenario, the objector simply feels that my account does not go far enough. I am not convinced that some sentences are vague at all finite orders, but I am also not convinced that no sentence is – and hence I wish to allow for all finite orders of vagueness. I do not, however, think that any sentence is $\omega$-order vague, and accordingly, my account does not make room for transfinite higher-order vagueness. The objector disagrees: she claims to have a genuine, robust intuition to the effect that some sentences are transfinite-order vague. Now if this is so, the objector can take up where I left off, and propose a semantics for vagueness in which the truth values are functions not just from finite sequences of elements of [0, 1] to [0, 1], but from transfinite sequences (or, of course, she might approach the matter in another way entirely). This theory might stop with sequences of a particular ordinal length, or it might countenance sequences of *every* ordinal length. The details would be complex – there might even be insuperable technical difficulties (or there might not). My objection to this approach, however, is simply that the added detail is pointless: *it is not in the service of any genuine intuition*.

In the second scenario, the objector feels that *no* account which assigns a unique truth value to a vague sentence could be adequate – not even if the truth value were a 'function' which assigned an element of [0, 1] to every 'function' from On (the class of all ordinals) to [0, 1]. But of course, in this form the objection is not really about higher-order vagueness at all: it is simply a version of the 'no non-vague theory of vagueness' objection, which was dismissed at the outset.

In sum, objector *A* asserts that there is nothing wrong with the very idea of assigning a unique truth value to a vague sentence – it's just that my proposed truth values are not good enough. This is the sort of objection that I made to the fuzzy theory. My response is that while the objection to the fuzzy theory is backed by genuine robust intuitions, the objection to my view is not. Objector *B* asserts that there is a problem with my proposed truth values, but ultimately the source of her conviction is an adherence to the thought that there is something wrong with the very idea of assigning a unique truth value to a vague sentence – and the more elaborate the truth values, the worse the problem. This is *not* the sort of objection that I made to the fuzzy theory. This sort of objection applies to *any* non-vague theory of vagueness – and in the absence of good arguments to the effect that no non-vague theory of vagueness can be correct, we may dismiss it as mere prejudice.[54]

## 12. The Sorites Paradox

Suppose we have a series of persons, ranging in height from four feet to seven feet, in increments of a thousandth of an inch. We number the persons from 0 through to 36,000, starting with the four foot person, assigning the next number to the next tallest person, and so on. '$Tx$' says that person number $x$ is tall, and '$xRy$' says that $x$ is the immediate predecessor of $y$ in the series. The Sorites argument is this:

1. $0R1$
2. $1R2$
3. $2R3$

⋮

36000. $35999R36000$
36001. $\neg T0$
36002. $\forall x \forall y((\neg Tx \wedge xRy) \rightarrow \neg Ty)$
36003. $\neg T36000$

I assume that premises 1–36001 are true to degree **T** and the conclusion (36003) is true to degree **F**.[55] The argument is classically valid, hence it is valid in blurry logic. That leaves premise 36002. This premise is *not* strictly greater than 0.5 true (to a first approximation). As we move along the series, facing an increasingly tall person $x$ at each stage, the degree of truth of the claim that $x$ is tall gradually increases. There may be a person $a$ such that '$Ta$' is 0.5 true (to a first approximation), or there may not (there may be adjacent persons $a$ and $b$ such that '$Ta$' is, say, 0.499 true and '$Tb$' is 0.501 true). Now premise 36002 is as true as its least-true instance. In either of the cases mentioned above, there is no instance less true than the one where the variable $x$ is replaced by a name of $a$, and the variable $y$ is replaced by a name of $a$'s successor: in the first case this instance is 0.5 true; in the second case it is 0.499 true. Either way, it is not strictly greater than 0.5 true. Hence the argument is unsound.

So far so good: we have located a flaw in the argument. But now what of the fact that the Sorites paradox is compelling? Premise 36002 is 0.5 true at best – so why is the argument as appealing as it is? Furthermore, it seems that premise 36002 is plausible because it corresponds to Closeness. Now one of my aims has been to accommodate Closeness – and yet here I am saying that the premise corresponding to Closeness is 0.5 true at best! Have I not got a big problem on my hands? No, I have not. The basic point is that although premise 36002 is plausible because it seems to correspond to Closeness, *it does not actually express Closeness*. We find it plausible because when read out, it sounds as though it is just a statement of Closeness – but closer attention reveals that this is not so.

If the predicate ‘is tall’ is vague, and if adjacent persons in the series are very similar in tall-relevant respects (i.e. in height), then Closeness tells us that for adjacent persons $a$ and $b$, ‘$Ta$’ and ‘$Tb$’ are very similar in respect of truth. In particular, if $a$ is not tall (at all), then it cannot be that $b$ is (definitely) tall. But isn’t this just what premise 36002 says? No! This is the crucial point. In the present semantics, $\neg Ta \rightarrow \neg Tb$ says the same thing as $Ta \vee \neg Tb$: ‘Either $a$ is tall, or $b$ is not’. Now when $a$ and $b$ are borderline cases for ‘tall’, ordinary speakers will have exactly the same sort of hedging reaction to this sentence as to the sentences ‘$a$ is tall’ and ‘$b$ is not tall’: they will not think that ‘Either $a$ is tall, or $b$ is not’ is clearly true.[56] Now the universal statement ‘For successive pairs $a$ and $b$, either $a$ is tall, or $b$ is not’ cannot be truer than its least-true instance: hence my semantics gives the correct result for premise 36002.

This still leaves the question as to why the Sorites paradox is compelling. In obtaining our assent to the inductive premise of her argument, the purveyor of the Sorites paradox will say, ‘Surely you do not think it can be the case that $a$ is *not tall*, and yet $b$ *is tall*?’ Given the emphases, it is natural to hear this as ‘Surely you do not think it can be the case that $a$ is clearly not tall, and yet $b$ is clearly tall?’ Now of course you don’t think this, so you say ‘No, of course not’, and things are under way. Now the purveyor of paradox decides to get tough, so she writes down an argument in formal language, and proves to you using your favourite proof theory for classical logic that it is valid. The premise she writes down corresponding to your rejection of the claim that $a$ is clearly not tall while $b$ is clearly tall is premise 36002. Now working within the classical picture, this is fine: saying that $a$ is tall is the same as saying that $a$ is fully or clearly tall, because in the classical picture, either you are tall or you are not – there are no halfway houses. But now we have moved to the blurry picture presented in this paper. The purveyor of paradox comes back and says ‘Look, this premise that you assented to so readily is only about half true on this new semantics – so the new semantics is wrong!’ But we need to be more subtle: we still vigorously reject the English statement that $a$ is clearly not tall while $b$ is clearly tall, *but given the new semantics, this claim is no longer accurately formalised as* $\neg Ta \rightarrow \neg Tb$. As we have just seen, the latter says something to which ordinary speakers make a hedging response. In the new framework, there *are* halfway houses between definite falsity and definite truth, and saying that *a* is tall is *not* the same as saying that *a* is clearly tall. The correct formalisation of the premise we heartily accept is $T_{\mathbf{F}}\mathbf{a} \rightarrow \neg T_{\mathbf{T}}\mathbf{b}$, where '**a**' is a name of the sentence '$Ta$' and '**b**' is a name of the sentence '$Tb$'. But with *this* premise, the purveyor of paradox is not going to be able to derive a paradoxical conclusion. Let '**x**' be a name for the claim 'Person $x$ is tall'; then '$T_{\mathbf{F}}\mathbf{0}$' says that the sentence 'Person 0 is tall' is true to degree **F**, and so on. Now consider the following argument:

1. $T_{\mathbf{F}}\mathbf{0} \rightarrow \neg T_{\mathbf{T}}\mathbf{1}$
2. $T_{\mathbf{F}}\mathbf{1} \rightarrow \neg T_{\mathbf{T}}\mathbf{2}$
3. $T_{\mathbf{F}}\mathbf{2} \rightarrow \neg T_{\mathbf{T}}\mathbf{3}$

$\vdots$

36000. $T_{\mathbf{F}}\mathbf{35999} \rightarrow \neg T_{\mathbf{T}}\mathbf{36000}$
36001. $T_{\mathbf{F}}\mathbf{0}$
36002. $\neg T_{\mathbf{T}}\mathbf{36000}$

Premises 1–36001 are all true to degree **T**. Conclusion 36002 is true to degree **F**. No problem: the argument is simply invalid.

In sum, my approach to the standard formulations of the Sorites paradox is that the arguments are valid, but the inductive premise (or premises) are not (all) sufficiently true to render the arguments sound. We nevertheless find the arguments plausible, because the inductive premise or premises are presented in such a way that we mistake them for expressions of Closeness. Closer inspection reveals that they are not so: interpreted in accordance with blurry semantics (as opposed to classical semantics) they express statements to which ordinary speakers would indeed make hedging responses; and when the expressions of Closeness are correctly formalised, paradoxical conclusions cannot be derived. Now this is quite different from the standard fuzzy approach. On the standard fuzzy approach, the semantics of the conditional is such that conditionals of the form $\neg Ta \rightarrow \neg Tb$ are true to ever so slightly less than degree 1: this is the proposed explanation of the appeal of the Sorites paradox. But even if this proposal provided a better explanation than mine of the attraction of the Sorites paradox as formulated in this section (and I do not believe that it does), the fuzzy theorist is left without anything to say about the Sorites paradox as formulated in other ways. Crispin Wright writes:

> Can a degree-theoretic account explain the plausibility of the major premises? There is no difficulty, of course, with the usual, quantified conditional form of premise. The explanation will claim that each instance, $Fa \rightarrow Fa'$, of $(\forall x)(Fx \rightarrow Fx')$ is *almost* true: that its consequent enjoys a degree of truth ever so nearly but not quite as great as that of its antecedent. And this claim will then be followed . . . by a stipulation that the degree of truth of any universally quantified statement is the minimum of the degrees of truth enjoyed by its instances . . . But . . . the major premise doesn't need to be conditional at all. In the case

of the Sorites-series of indiscriminable color patches for instance, we could just as well take it in the form

$$(\forall x) - [\text{red}(x) \ \& \ - \text{red}(x')].$$

All the ways of making the conditional form of major premise seem intuitively plausible would be applicable to this conjunctive form ... [the degree theorist] needs to explain ... with what right such a conjunctive major premise may be regarded as *almost true*; otherwise he cannot explain its plausibility, or duly acknowledge the force of the arguments which seem to sustain it. (Wright, 1987, pp. 251–252)

As Wright then goes on to point out, one cannot see how the degree theorist could give an account on which such a conjunctive major premise is almost true: and in any case, on the standard fuzzy account, such premises are not almost true. Thus, the standard fuzzy theorist's explanation of the plausibility of the conditional formulation of the Sorites paradox does not extend to other formulations – whereas my explanation does. Of course, the fuzzy theorist – having the resources to accommodate Closeness – can adopt the sort of view I have advocated: my point is precisely that the fuzzy theorist should adopt my explanation of the plausibility of Sorites paradoxes, not vice versa.

## 13. Semantic Indeterminacy and Modality

In (Smith, 2001) I argue that in the standard cases of vagueness, what we need is a theory that accounts for vagueness in language in terms of vagueness in the world, rather than seeing it as a purely semantic phenomenon.[57] But while the idea of semantic indeterminacy should not play the *central* role in the theory of vagueness, we should allow for its *possibility*. In fact it is a very straightforward matter to accommodate semantic indeterminacy within an extension of the view presented in this paper. The extension runs along supervaluationist lines.

First, we will want to allow that in some cases there might be *more than one* admissible interpretation of a discourse. When I say 'Bob is bald', it might be that there is no unique blurry set that is the extension of my predicate 'is bald': rather, my predicate picks out a number of blurry sets, each of which meets all the constraints on the correct or intended interpretation.[58] Now in the standard supervaluationist view, a sentence is true simpliciter if it is classically true on *all* admissible interpretations. So to determine how true my sentence 'Bob is bald' is, we should, in effect, quantify over the admissible interpretations. Thus if $\{\mathfrak{M}_i\}$ is the set of all admissible interpretations of my utterance, and $f_i$ is the truth value of my utterance on $\mathfrak{M}_i$, then the supertruth value of my utterance with respect to the supervaluation $\{\mathfrak{M}_i\}$ is $\bigwedge\{f_i\}$.

The point of this extension of the framework presented in this paper is somewhat different from the point of the standard supervaluationist account of vagueness. Vagueness has already been accommodated within the framework in which each discourse has a unique correct blurry interpretation; the extension is to allow for the possibility that in some cases, the facts that fix which interpretation of our sentences is the correct one do not suffice to single out a unique such interpretation, but rather leave open a number of possibilities. In the standard supervaluationist account of vagueness, the admissible interpretations are the classical interpretations which precisify the actual incomplete interpretation, while respecting the constraints of penumbral connection. In the present account, the admissible interpretations will simply be those which are not ruled out by the facts that fix which interpretations of our sentences are correct – whatever these facts are.

Having made room for the idea of multiple correct interpretations, it is easy to see (in outline) how to develop blurry modal logics. Suppose that we have a set $W$ of worlds – each associated with an interpretation of our language – and an accessibility relation $\leq$ on this set. If $[\mathcal{A}]_w$ is the truth value (i.e. degree function) of sentence $\mathcal{A}$ at world $w$, then

$$[\Box\mathcal{A}]_w = \bigwedge\{[\mathcal{A}]_v : w \leq v\},$$

$$[\Diamond\mathcal{A}]_w = \bigvee\{[\mathcal{A}]_v : w \leq v\}.$$

## 14. Comparison with Other Views

One usually finds a fuzzy set defined as a function from some universal set to the closed interval [0, 1]. However, [0, 1]-valued fuzzy sets are in fact just one sort (the basic sort) of fuzzy set: following the terminology of (Klir and Yuan, 1995, p. 16), they are *ordinary fuzzy sets*. There are many generalisations of and variations on ordinary fuzzy sets; I shall end this paper by considering some views that are (apparently or in fact) related to the view presented here.

An ordinary fuzzy subset $S$ of a universal set $U$ has a membership function of the form:

$$S : U \to [0, 1].$$

Let $\mathfrak{F}[0, 1]$ be the set of all fuzzy subsets – or fuzzy power set – of [0, 1], i.e. the set $[0, 1]^{[0,1]}$ of all functions $S : [0, 1] \to [0, 1]$. A *type* 2 *fuzzy subset* $S$ of a universal set $U$ has a membership function of the form:

$$S : U \to \mathfrak{F}[0, 1] \quad \text{or equivalently} \quad S : U \to [0, 1]^{[0,1]}.$$

Thus where an ordinary fuzzy subset of $U$ assigns a number in [0, 1] to each element of $U$, a type 2 fuzzy subset assigns an ordinary fuzzy subset of [0, 1] to each element of $U$. Let $\mathfrak{F}^2[0, 1]$ be the set of all type 2 fuzzy subsets – or type 2 fuzzy power set – of [0, 1], i.e. the set $([0, 1]^{[0,1]})^{[0,1]}$ of all functions $S : [0, 1] \to [0, 1]^{[0,1]}$. A *type* 3 *fuzzy subset* $S$ of a universal set $U$ has a membership function of the form:

$$S : U \to \mathfrak{F}^2[0, 1] \quad \text{or equivalently} \quad S : U \to ([0, 1]^{[0,1]})^{[0,1]}.$$

Thus where an ordinary fuzzy subset of $U$ assigns a number in [0, 1] to each element of $U$, and a type 2 fuzzy subset assigns an ordinary fuzzy subset of [0, 1] to each element of $U$, a type 3 fuzzy subset assigns a type 2 fuzzy subset of [0, 1] to each element of $U$. The definitions of type $n$ fuzzy subset and type $n$ fuzzy power set proceed in the obvious recursive way.[59]

There are some points of contact between my view and type $n$ fuzzy set theory, but overall, similarities are outweighed by differences. A blurry subset is not a type $n$ fuzzy subset, for any finite $n$. A blurry subset of $U$ assigns a *DF* to each member of $U$, where a *DF* in turn assigns an element of [0, 1] to each finite sequence of elements of [0, 1]. A type $n$ fuzzy subset of $U$ assigns to each member of $U$ a type $(n-1)$ fuzzy subset of [0, 1], which in turn may be represented (although this is not normal practice) as a function from length $(n-1)$ sequences of elements of [0, 1] to [0, 1]. Thus, a blurry subset is not even a type $(\omega + 1)$ fuzzy subset, this being thought of as assigning to each element of $U$ a function from $\omega$-length sequences of elements of [0, 1] to [0, 1] – for a *DF* assigns a value to finite sequences of every length, rather than only to sequences of infinite length. We *could*, however, think of a blurry subset of $U$ as an infinite sequence $(S^1, S^2, S^3, \ldots)$ of a type 1 fuzzy subset of $U$, a type 2 fuzzy subset of $U$, a type 3 fuzzy subset of $U$, and so on. However, not just *any* such sequence will count as a blurry set: for as yet we have not mentioned the idea of *approximation*. In a degree function, the assignments to longer sequences provide a more detailed approximation (of degree of membership or truth) than the assignments to shorter sequences: specifically, the link is that the assignments to the sequences $\langle a_1, \ldots, a_n, x\rangle$ encode a density function of which the assignment to $\langle a_1, \ldots, a_n\rangle$ is the centre of mass. Thus, if we are to think of a blurry set as a sequence of type $n$ fuzzy sets, we will need to impose some restrictions on these fuzzy sets: *for example*, for any $a \in U$, we will want it to be the case that $\int_0^1 x(\mathfrak{f} \circ S^2(a))(x)\,\mathrm{d}x = S^1(a)$.

This is in fact the crucial difference between blurry sets and fuzzy sets of higher type: I distinguish actual degrees of membership or truth (i.e. degree functions) from approximations thereto (i.e. fuzzy degrees), whereas there is no such conceptual distinction in the higher-type fuzzy view. The fact that I draw this distinction makes the question of defining intersection, union and so on for blurry sets (equivalently, of defining the connectives and so on in blurry semantics) more difficult than the corresponding question for type $n$ fuzzy sets, but on the other hand, it gives my view a conceptual coherence that the higher-type fuzzy view lacks. I shall now explain these two claims in more detail.

The set-theoretic operations for fuzzy sets of higher type are defined using the *extension principle*.[60] Consider a function $f : X_1 \times \cdots \times X_r \rightarrow Y$, with $y = f(x_1, \ldots, x_r)$, where $Y$ and the $X_i$ are ordinary crisp sets. Given $r$ ordinary fuzzy sets $S_i$ on each of the $X_i$, the extension principle yields a fuzzy set $S$ on $Y$, through $f$:

$$\begin{aligned} S(y) &= \sup_{x_1,\ldots,x_r : y = f(x_1,\ldots,x_r)} \min(S_1(x_1), \ldots, S_r(x_r)) \\ &= 0 \text{ if } f^{-1}(y) = \emptyset. \end{aligned}$$

Consider a particular case, say conjunction/intersection: for ordinary fuzzy sets we have an operation $\wedge : [0, 1] \times [0, 1] \rightarrow [0, 1]$; for type 2 fuzzy sets we need an operation $\wedge : [0, 1]^{[0,1]} \times [0, 1]^{[0,1]} \rightarrow [0, 1]^{[0,1]}$; and given the former, the extension principle yields the latter. Given operations for fuzzy sets of type 2, the extension principle again yields operations for fuzzy sets of type 3, and so on.[61]

If we wanted to apply this method of defining operations to the case of blurry sets, we would proceed as follows: thinking of a blurry subset $S_i$ of $U$ as an infinite sequence $(S_i^1, S_i^2, S_i^3, \ldots)$ of a type 1 fuzzy subset of $U$, a type 2 fuzzy subset of $U$, a type 3 fuzzy subset of $U$, and so on, we would define (for example) $S_1 \cap S_2$ as $(S_1^1 \cap S_2^1, S_1^2 \cap S_2^2, S_1^3 \cap S_2^3, \ldots)$, where the $S_i^j \cap S_k^j$ are defined using the extension principle. However, this definition simply does not work! For some particular object $a \in U$, and two arbitrary blurry subsets $S_1$ and $S_2$ of $U$, consider the two degree functions $S_1(a)$ and $S_2(a)$. I mentioned above that we need to impose some restrictions on the $S_i^j$ if we are to think of $S_i$ as a sequence $(S_i^1, S_i^2, S_i^3, \ldots)$ of higher-type fuzzy sets. Let us suppose that we have imposed the appropriate restrictions. We now want it to be the case that if $S_1$ and $S_2$ meet the restrictions, then so does $S_1 \cap S_2$ as just defined. Thus, for example, we have it that $\int_0^1 x(\mathfrak{f} \circ S_1^2(a))(x)\,\mathrm{d}x = S_1^1(a)$ and $\int_0^1 x(\mathfrak{f} \circ S_2^2(a))(x)\,\mathrm{d}x = S_2^1(a)$, and we want it to be the case that $\int_0^1 x(\mathfrak{f} \circ (S_1^2 \cap S_2^2)(a))(x)\,\mathrm{d}x = (S_1^1 \cap S_2^1)(a)$. The problem is that in general this will not be the case: the definition of intersection just presented (and the same goes for the analogous definitions of union and so on) does *not* respect the new restrictions. For example, if there is no $x \in [0, 1]$ such that *both* $S_1^2(a)(x) > \epsilon$ and $S_2^2(a)(x) > \epsilon$, for some sufficiently small $\epsilon$, then $(S_1^2 \cap S_2^2)(a)(x) \leq \epsilon$ for every $x \in [0, 1]$, and hence does not encode a normalised density function at all, let alone one whose centre of mass is at $(S_1^1 \cap S_2^1)(a)$.

Summing up: with a bit of rejigging (viz., thinking of a blurry set as a sequence of type *n* fuzzy sets) we can find some common ground between my view and higher-type fuzzy views.[62] This common ground is very limited, however: when we come to the issue of connectives and set-theoretic operations, the views diverge significantly. This divergence can be traced ultimately to the distinction in my view between actual degrees of membership/truth, and approximations thereto. Now even if this distinction did not lead to a divergence at the level of formal details, it would still be very significant in itself – for the chief problem with type *n* fuzzy sets is that they are not sufficiently motivated or explained at the conceptual level. Before discussing this second point, however, I shall introduce a particular type 2 fuzzy view, due to Zadeh: the theory of linguistic truth values.[63] This theory faces a problem in connection with the definition of the connectives that is very similar to the problem discussed in the previous paragraph, and it also provides a clear example of the conceptual problems associated with higher-type fuzzy views.

In accordance with (reasonably) standard terminology, I have described [0, 1]-valued logic as *fuzzy logic*, and $[0, 1]^{[0,1]}$-valued logic as *type* 2 *fuzzy logic*.[64] Zadeh uses a different terminology, according to which [0, 1]-valued logic is *nonfuzzy*, and the term ‘fuzzy logic’ is reserved for the generalisation that Zadeh introduces: “A fuzzy logic, FL, may be viewed, in part, as a fuzzy extension of a nonfuzzy multi-valued logic which constitutes a *base* logic for FL. For our purposes, it will be convenient to use as a base logic for FL the standard Łukasiewicz logic $L_1$ (abbreviated from $L_{\text{Aleph}_1}$) in which the truth-values are real numbers in the interval [0, 1]” (Zadeh, 1975b, pp. 409–410). The truth-value set of FL is a *countable* set $\mathcal{T}$ of the form

> {*true*, *false*, *not true*, *very true*, *not very true*, *more or less true*, *rather true*, *not very true and not very false*, ...}

where each element of this set represents a fuzzy subset of the truth-value set of $L_1$, that is, of [0, 1].[65]

Given the standard operations of [0, 1]-valued logic, Zadeh defines logical operations in FL by means of the extension principle, as explained above. Here he runs into a problem similar to that which I would face were I to think of blurry sets as sequences of type *n* fuzzy sets, and then define set-theoretic operations using the extension principle. The problem for Zadeh is that if we start with two FL truth values $\phi$ and $\psi$ (that is, two fuzzy subsets of [0, 1] *that are named by terms in* $\mathcal{T}$) and apply the definition of (say) conjunction that the extension principle yields, we will end up with a fuzzy subset of [0, 1], *but it might not be named by any term in* $\mathcal{T}$, in which case it is *not* an FL truth value.[66] Zadeh's solution involves the idea of *linguistic approximation*: if (for example) $\phi \wedge \psi$, as defined using the extension principle, is not named by any member of $\mathcal{T}$, then we *approximate* it by a linguistic truth value, that is, by a fuzzy subset of [0, 1] that *is* named by a member of $\mathcal{T}$. Zadeh notes that in general there will not be a unique linguistic approximation. This strikes me as serious: it means that even if sentences '*A*' and '*B*' have unique truth values, '*A* and *B*' might not have a unique truth value (which is quite different from its having an unknown truth value).

Let us now consider the conceptual issues surrounding higher-type fuzzy sets. At first sight, it seems that many generalisations of ordinary fuzzy logic are motivated by thoughts similar to those that motivated me: as Klir and Yuan sum up the situation, "The primary reason for generalizing ordinary fuzzy sets is that their membership functions are often overly precise. They require that each element of the universal set be assigned a particular real number" (Klir and Yuan, 1995, p. 16).[67] However this impression does not survive closer scrutiny; for example, the passage just quoted continues: "However, for some concepts and contexts in which they are applied, we may be able to identify appropriate membership functions only approximately" (Klir and Yuan, 1995, p. 16). Such considerations about the *identification* of membership functions are insufficient to motivate a view like mine: rather, they motivate *epistemicism* (built on fuzzy, rather than classical foundations). Again, consider the following quotation, from the first page of a recent article on type 2 fuzzy sets: "Type-2 fuzzy sets allow us to handle linguistic uncertainties, as typified by the adage "words can mean different things to different people" " (Karnik et al., 1999). Such considerations about variations in meaning from person to person are likewise insufficient to motivate a view like mine: rather, they motivate *contextualism* (built on fuzzy, rather than classical or three-valued foundations). In general, most writers on generalised fuzzy sets provide just a line or two concerning motivation: sometimes the phrases used suggest epistemic worries, sometimes worries about context, and sometimes just a general worry about the restrictiveness of ordinary fuzzy logic. It is, I believe, fair to sum up the basic motivation of most writers on higher-type fuzzy sets as follows: "In applications, we are not presented with a single real number to serve as $x$'s degree of $X$: the data are more complex. We can massage the data to fit the ordinary fuzzy theory, but it would be better to have a more complex theory, with more sockets into which to plug data." Now most of these writers are engineers and computer scientists, so this is fine for their purposes and audiences – but it would not be fine in a philosophical treatment of vagueness. Thus, to at least as great an extent as the formal differences, what distinguishes my view from existing generalisations of ordinary fuzzy sets is the effort I have expended to ensure that the details of my view are well motivated and *conceptually* coherent.

This point is most significant when it comes to comparing my view with the views of those writers on generalised fuzzy sets who *do* have philosophical goals. Two such writers are Zadeh and Copeland. Zadeh writes: "approximate reasoning deals, for the most part, with propositions which are fuzzy rather than precise, e.g., 'Vera is *highly intelligent*', 'Douglas is *very inventive*', 'Berkeley is *close* to San Francisco', 'It is *very likely* that Jean-Paul will *succeed*', etc. Clearly, the fuzzy truth-values of FL are more commensurate with the fuzziness of such propositions than the numerical truth-values of $L_1$" (Zadeh, 1975b, p. 416). I think I probably share the feeling expressed in the last sentence – but it is difficult to say, because the relationship between the fuzzy truth values of FL and the numerical truth values of $L_1$ is never elucidated: all we are told is, "The truth-values in $\mathcal{T}$ are referred to as *linguistic* truth-values in order to differentiate them from the numerical truth-values of $L_1$" (Zadeh, 1975b, p. 412). We certainly have a distinction here, but what is the *difference* which it marks? Formally, FL is parasitic on $L_1$, and yet *conceptually*, the relationship between them is simply never explained. This is in contrast to my own view, where fuzzy degrees are taken to be *approximations* to actual degrees (i.e. degree functions).[68]

Copeland discusses a version of higher-type fuzzy logic in which:

> With each statement $A$ is associated $A$'s *higher-order profile*, written $\Omega(A)$. $\Omega(A)$ is a subset of $[0, 1] \times [0, 1]$ . . . The idea is that each member $\langle n, i\rangle$ of $\Omega(A)$ records the degree $i$ to which it is true that $A$ possesses the value $n$. I shall call the right-hand members of pairs in a profile *higher-order degrees* and the left-hand members *primary degrees* . . . Where $A$ definitely has the value $n$, then $\langle n, 1\rangle \in \Omega(A)$ . . .
>
> Each $\Omega(A)$ considered so far consists of ordered pairs. Higher-order profiles of this sort are called *two-dimensional* . . . A three-dimensional higher-order profile consists of triples $\langle n, i, j\rangle$. $j$ is the degree to which it is true that $i$ is the degree to which it is true that the sentence in question possesses the value $n$. This process can be continued indefinitely . . . It is, of course, only the finite cases that are of use to engineers, and how low a dimensionality one can get away with will depend on the software project in hand. (Copeland, 1997, pp. 531–532)[69]

Despite any impression that might be given by the last sentence just quoted, Copeland's article is primarily a contribution to philosophy (and indeed appeared in *The Journal of Philosophy*). Thus it would not be a sufficient defence of his view to say simply that it gives us more sockets into which to plug numbers. Now Copeland distinguishes *higher-order degrees* and *primary degrees*: but as in Zadeh's case, all we have here is a distinction, without any elucidation of the underlying difference. And indeed, if we consider the issue carefully, we can see that while there is no formal problem with Copeland's view, the view does appear to be *conceptually* inadequate. Consider the initial, two-dimensional view. The idea might be that each sentence $A$ has a fuzzy degree of truth (a number in [0, 1]) as well as a higher-order profile (i.e. the ordinary fuzzy theory is *augmented*), or the idea might be that each sentence $A$ just has a higher-order profile (i.e. the ordinary fuzzy theory is *replaced*). Textual evidence suggests that the latter idea is intended, but I am not quite sure, so I shall consider both options. In the second case, sentence $A$ has a higher-order profile, that is, a set of ordered pairs of elements of [0, 1]. If one of these pairs is, say, $\langle 0.5, 0.8\rangle$, then the idea is supposed to be that the degree to which $A$ possesses the value 0.5 is 0.8. Now immediately we have a problem: for sentences do not get assigned fuzzy values, they only have higher-order profiles; hence the degree to which $A$ possesses the value $x$ is 0, for all $x$ in [0, 1]. We cannot say that sentences get assigned higher-order profiles, rather than fuzzy degrees, and then also say that there is some positive degree of truth to the statement that some sentence has some fuzzy degree! So what about the first case, in which sentences have fuzzy degrees as well as higher-order profiles? Well, the situation here is no better. Suppose that $A$'s fuzzy degree is $d$; then something has gone very wrong if $A$'s higher-order profile does not consist just of the pairs $\langle d, 1\rangle$ and $\langle x, 0\rangle$, for every $x \neq d$. For one cannot say that $A$'s fuzzy degree is $d$ (rather than $x$, for any $x \neq d$), and then also say that it is true to some positive degree that $A$'s fuzzy degree is $x$, for some $x \neq d$!

My view does not face this conceptual problem, because I am careful to distinguish the roles played in my account by fuzzy degrees and degree functions. The actual truth values that sentences have are degree functions; in general these are unknown to speakers; they may however be approximated by assigning fuzzy degrees to sentences. If I say that some sentence is true to degree 0.5, it is understood that this is an approximation to its actual truth value; if you say that my statement is true to degree 0.9, this provides (part of) a more detailed approximation; and so on. Once we have this idea, we need some constraints governing the relationships between successive levels of approximation, in order to make good on the idea that subsequent approximations really are better, or more detailed, than earlier approximations; this is where the idea of density functions and centres of mass comes in. This leads to greater *formal* complexity in my view, as compared with type *n* fuzzy views – but this is a small price to pay for *conceptual* coherence, which is something that type *n* fuzzy views lack.

## 15. CONCLUSION

In this paper I have presented a theory of vagueness with the following features. First, like the fuzzy theory, my theory accommodates vagueness as characterised in terms of Closeness, but my theory also accommodates the higher-order vagueness that the fuzzy account cannot accommodate. Second, by carefully distinguishing actual degrees of membership/truth (i.e. degree functions) from approximations thereto (i.e. fuzzy degrees), I am able to accommodate the intuition that for any statement $S_1$ to the effect that 'Bob is bald' is $x$ true, for $x$ in [0, 1], there should be a *further* statement $S_2$ which tells us how true $S_1$ is, and so on – that is, I am able to accommodate higher-order vagueness – *without* resorting to the claim that the language in which the semantics of vagueness is presented is itself vague: rather than a hierarchy of assignments of simple fuzzy truth values, I employ a single assignment of complex truth values which have an internal hierarchical structure. Third, my theory does not require us to abandon the idea that the logic – as opposed to the semantics – of vague discourse is classical.

## ACKNOWLEDGEMENTS

I am very grateful to the following: John P. Burgess and Gideon Rosen, who supervised my PhD dissertation, on parts of which this paper is based; Amitavo Islam and Ed Mares, with each of whom I have had many enjoyable and fruitful discussions of this material; Dick Jeffrey, who read several versions and gave me extensive feedback; Graham Priest and Greg Restall, who gave me detailed comments on an early version in 1999; and Bob Brandom, Kit Fine, Mark Johnston, Scott Soames, Steve Weinstein, an anonymous referee for this Journal, and audiences at Princeton University (1999), at the University of Sydney (2000), at the Annual Conference of the Australasian Association of Logic (1999), and at the Annual Conference of the Australasian Association of Philosophy (1999 and 2000), for helpful comments and discussions.

## NOTES

[1] See for example (Williamson, 1994, p. 127) and (Copeland, 1997, p. 522).

[2] I am not drawing an analogy between the *content* of my view and the content of Mandelbrot's view; I am drawing an analogy between the *structure* of Mandelbrot's criticism of the idea that coastlines are Euclidean curves and the structure of my criticism of the fuzzy account of vagueness. For my purposes, it is quite irrelevant whether Mandelbrot's view is correct.

[3] I shall sometimes distinguish a set from its characteristic function, and sometimes identify them – but I trust this will never cause any confusion.

[4] A model theory for a standard first-order language, that is – the idea being that what distinguishes those parts of language which are vague from those parts which are precise is a matter not of syntax, but of semantics.

[5] At this stage, there is no requirement that the curve have a nice shape, as it does in Figure 2.

[6] The intuitive idea here is as follows. Imagine taking the unit interval [0, 1] and bending it into an arc which constitutes the lower-right quarter of a circle centred on the point $a$. Now draw a straight line from $a$ through some point $x$ on the arc (i.e. some point $x$ in [0, 1]) and down to the horizontal axis. The point on the horizontal axis which this line hits is the value to which $\mathfrak{f}$ maps the point $x$.

[7] The general distinction between masses and densities also crops up in probability theory, in the contrast between discrete and continuous probabilities. For another illustration of the general distinction, suppose one has a number of lead fishing sinkers on a wire. The sum of all their masses is 1, and the length of the wire is also 1. I can describe the distribution of mass on the wire by assigning a number to each point on the wire: if the point is right at the centre of a sinker, this number is the mass of that sinker, and otherwise it is 0. The sum of all these numbers will be 1. But suppose that I now heat all the sinkers so that they melt and flow together along the wire. Then I can again describe the distribution of mass on the wire by assigning a number to each point on the wire: but this time the numbers will represent densities, not masses, and they will not sum to 1. These numbers together determine a density function $f(x)$: the mass on any section $[a, b]$ of the wire is $\int_a^b f(x)\,\mathrm{d}x$, and because the total mass of the sinkers was 1, $\int_0^1 f(x)\,\mathrm{d}x = 1$.

[8] It is worth mentioning at this point that instead of bringing in $\mathfrak{f}$ in order to maintain the idea that the area under each density function is 1, we could leave $\mathfrak{f}$ out of the picture and just consider $f$, but stipulate that the 'unit' area under each density function is not 1 but some smaller number, say 0.5. This would make no essential difference to my view, although it would make the *presentation* of the material in Section 4 less straightforward.

[9] Thus, which members of $F$ count as Type I (Basic) *DF*'s depends upon our choice of $\mathfrak{f}$ – but obviously, if a *DF* is to encode something, then whether or not it does so is relative to the choice of coding method.

[10] In general, the centre of mass of a density function $f(x)$ over an interval $[a, b]$ is $\int_a^b xf(x)\,\mathrm{d}x / \int_a^b f(x)\,\mathrm{d}x$. Where – as in our case – the denominator is equal to 1, this reduces to the numerator alone.

[11] In these definitions, $n \geq 0$. I adopt the terminological convention that in the case $n = 0$, $\langle x_1, \ldots, x_n \rangle = \langle\rangle$ and $\langle x_1, \ldots, x_n, k \rangle = \langle k \rangle$.

[12] $\mathfrak{c}$ is the cardinality of the set of real numbers.

[13] Each Type I (Continuous) *DF* is determined by a choice of a number in [0, 1], followed by a choice of a continuous function $f^n$ from the $n$-dimensional unit cube to [0, 1] for each positive integer $n$. In each case there are $\mathfrak{c}$ things to choose from, so there are $\mathfrak{c}$ Type I (Continuous) *DF*'s. There are 2 Type II *DF*'s. For each Type I (Continuous) *DF* $f$ there are $\aleph_0$ (i.e. natural-number many) corresponding Type III *DF*'s $f_n$, one for each non-negative integer $n$. Thus there are $\mathfrak{c}$ Type III *DF*'s. Thus there are $\mathfrak{c} + 2 + \mathfrak{c} = \mathfrak{c}$ *DF*'s in total.

[14] The first approach is carried through in detail in (Smith, 2001, pp. 189ff.).

[15] Note that this terminology concerns vagueness of *sentences*, not of *predicates*.

[16] If there were such a measure: at this stage I have not said anything about how spread might be measured. At the moment we are working at a fairly rough, intuitive level; things will be made more precise below.

[17] For a detailed discussion of the distinction between semantic and worldly predicate vagueness, see (Smith, 2001) and (Rosen and Smith, 2004). Briefly, the idea is that a predicate '*F*' is semantically vague if it fails to pick out a unique property, but rather refers ambiguously to many properties at once, each of these properties being sharp – that is, such that any object either possesses the property *simpliciter* or fails to possess it *simpliciter*. A predicate '*F*' is vague in the worldly sense if it picks out a unique property, this property being inherently vague – that is, such that objects need not possess it *simpliciter* or fail to possess it *simpliciter*, but may possess it to intermediate degrees. The standard supervaluationist account, for example, sees vagueness as a semantic phenomenon, whereas the standard fuzzy account, for example, sees vagueness as a worldly phenomenon.

[18] Distinguish a *normal* density function (which I am about to explain) from a *normalised* density function (the area under the graph of which is equal to 1, as explained above).

[19] Fabio has lots of hair and Yul Brynner has none at all.

[20] Note carefully that while a normal density function assigns values to all real numbers, each $\mathfrak{f} \circ f_{\langle a_1,\ldots,a_n\rangle}$ is defined only on [0, 1]. I reiterate this point in order to ward off a particular misunderstanding. An anonymous referee noted that the normal distribution is symmetric about its mean, and inferred that if a sentence is 0.1 true to a first approximation, then on my view it must be the case that at the second level of approximation, both 0.25 and $-0.05$ are regarded as equally good first approximations, which is absurd. I agree that this result is absurd: but it is not a consequence of my view. For while $N(\mathfrak{f} \circ f_{\langle\rangle})$ must assign the same value to $-0.05$ as it assigns to 0.25, $\mathfrak{f} \circ f_{\langle\rangle}$ does not assign a value to $-0.05$ at all. $N(\mathfrak{f} \circ f_{\langle\rangle})$ and $\mathfrak{f} \circ f_{\langle\rangle}$ are extremely similar, but they are not identical: recall that $\int_0^1 (\mathfrak{f} \circ f_{\langle\rangle})(x)\,\mathrm{d}x = 1$ while $\int_0^1 (N(\mathfrak{f} \circ f_{\langle\rangle}))(x)\,\mathrm{d}x < 1$.

[21] We could not (in general) require that they have the same standard deviation; this is why we take $\pi$, rather than $\sigma$, as our measure of spread.

[22] Note that we are simply *representing* or *naming DF*'s using sequences of real numbers: we are not *reducing DF*'s to sequences of real numbers – *DF*'s are (still) certain functions from $[0, 1]^*$ to [0, 1]; all we are doing now is introducing a convenient way of referring to each *DF*.

[23] Going in the other direction, $f$ can be completely reconstructed from its corresponding sequence. The first element of the sequence gives the value assigned to $\langle\rangle$, which is the mean of the density function which $f_{\langle\rangle}$ encodes. The second element gives the precision of this function, from which we can calculate its standard deviation, hence the function itself (here, for the sake of simplicity of presentation, I conflate $\mathfrak{f} \circ f_{\langle a_1,\ldots,a_j\rangle}$ and $N(\mathfrak{f} \circ f_{\langle a_1,\ldots,a_j\rangle})$). The values of (the encoded version of) this function give the means of

the density functions at the next level, and the third member of the sequence gives their precision, hence their standard deviations – and so on all the way up.

[24] Note carefully what has just been done: first we named *DF*'s using sequences of reals; then we defined an ordering of *DF*'s via an ordering of their names (compare asking school children to line up in alphabetical order: one transfers the alphabetical ordering of their names onto the students themselves). One consequence of this ordering is that where $C$ is a sentence with truth value represented by the sequence $\langle 0.1, 0.9, 0.9, 0.9, \ldots \rangle$ and $D$ is a sentence with truth value represented by the sequence $\langle 0.1, 0.1, 0.1, \ldots \rangle$, $C$ is truer than $D$, because although $C$ is precisely as true as *D at the first level of approximation*, $C$'s degree of truth is, *overall*, less diffuse. An anonymous referee objected here that this lack of diffuseness should not render *C truer* than $D$, as "what we are getting precise about is just how *false C* really is". This objection reveals an important misunderstanding, to dispel which is the purpose of this note. The correct way to interpret $C$'s *DF* $\langle 0.1, 0.9, 0.9, 0.9, \ldots \rangle$ is *not* as a *truth* value (0.1) followed by a hierarchy of *vagueness* values. 0.1 is not $C$'s truth value! 0.1 is simply a *first approximation* to $C$'s degree of truth. First approximations should not be given a privileged role when it comes to determining the ordering of *DF*'s: higher-level approximations (assignments to sequences of length 1 and greater) are just as much parts of a *DF* as the first approximation (the assignment to the empty sequence), and should be given just as big a role when it comes to ordering the *DF*'s. The entire *DF* is $C$'s *truth* value – not just the assignment it makes to the empty sequence. To think otherwise is to abandon my view in favour of another sort of generalisation of fuzzy logic: a sort which I shall criticise in Section 14.

[25] We shall see in Section 7 that where $f$ is the truth value of the sentence $C$, $f'$ is the truth value of the sentence $\neg C$. Given this, it follows from the definition of $'$ just given that if the degree of truth of $C$ is very diffuse, then the degree of truth of $\neg C$ is very localised. An anonymous referee objected here that this seems wrong, and suggested that we should define the negation of $\langle f_1, f_2, f_3, \ldots \rangle$ as $\langle f_1', f_2, f_3, \ldots \rangle$. I disagree. For a start, given the latter definition, it would not be the case that if $f \leq g$, then $g' \leq f'$, which is very undesirable. More importantly, the objection once again (cf. footnote 24) privileges the first level of approximation: the negation of $C$ should be bad in respect of truth in *every* way in which $C$ is good in respect of truth – and while having a low first approximation is indeed bad, *so is being diffuse*.

[26] The two conditions for an involution are that for all $f$ and $g$ in *DF*, (i) $(f')' = f$ and (ii) $f \leq g \Rightarrow g' \leq f'$. (i) is just 14. Note that in (ii), with regard to the ordering $\leq$, $f \leq g \Leftrightarrow f \wedge g = f \Leftrightarrow f \vee g = g$. So we show that $f \vee g = g \Rightarrow g' \wedge f' = g'$. If $f \vee g = g$ then $(f \vee g)' = g'$; by 12, $(f \vee g)' = f' \wedge g'$, and by 5, $f' \wedge g' = g' \wedge f'$, so $g' \wedge f' = g'$.

[27] This is just the standard notion of identity for functions.

[28] *Proof.* (Condition 1:) For arbitrary $u \in U$, $S(u) = f \in DF$. Either $f = \mathbf{F}$, or $f \neq \mathbf{F}$. Former case: $S^*(u) = \mathbf{T}$. $\mathbf{F} \wedge \mathbf{T} = \mathbf{F}$. So $(S \cap S^*)(u) = \mathbf{F}$. Latter case: $S^*(u) = \mathbf{F}$. For any $f \in DF$, $\mathbf{F} \wedge f = \mathbf{F}$. So $(S \cap S^*)(u) = \mathbf{F}$. Thus $\forall u \in U$, $(S \cap S^*)(u) = \mathbf{F}$, i.e. $S \cap S^* = \emptyset_v$. (Condition 2:) We need to show that $S_2 \cap S_1^* = S_2 \Leftrightarrow S_1 \cap S_2 = \emptyset_v$. ($\Rightarrow$:) $\forall u \in U$, $(S_2 \cap S_1^*)(u) = S_2(u) \wedge S_1^*(u) = S_2(u)$. Case (i): $S_1^*(u) = \mathbf{F}$. Then $S_2(u) = \mathbf{F}$ also, because $S_2(u) \wedge S_1^*(u) = S_2(u)$. Thus $S_1(u) \wedge S_2(u) = \mathbf{F}$. Case (ii): $S_1^*(u) = \mathbf{T}$. Then $S_1(u) = \mathbf{F}$. Thus $S_1(u) \wedge S_2(u) = \mathbf{F}$. So in all cases, $S_1(u) \wedge S_2(u) = \mathbf{F}$. So in all cases, $(S_1 \cap S_2)(u) = \mathbf{F}$; i.e. $S_1 \cap S_2 = \emptyset_v$. ($\Leftarrow$:) $\forall u \in U$, $(S_1 \cap S_2)(u) = S_1(u) \wedge S_2(u) = \mathbf{F}$. Case (i): $S_1(u) = \mathbf{F}$. Then $S_1^*(u) = \mathbf{T}$, hence $S_2(u) \wedge S_1^*(u) = S_2(u)$. Case (ii): $S_1(u) \neq \mathbf{F}$. Then

$S_1^*(u) = \mathbf{F}$, hence $S_2(u) \wedge S_1^*(u) = \mathbf{F}$. But also $S_2(u) = \mathbf{F}$, because $S_1(u) \wedge S_2(u) = \mathbf{F}$ and $S_1(u) \neq \mathbf{F}$. So in all cases, $S_2(u) \wedge S_1^*(u) = S_2(u)$, hence $S_2 \cap S_1^* = S_2$.

29 Note that the treatment of the case $f_x(\langle\rangle) = 0.5$ is arbitrary.

30 Note that the result does not hold if we replace $< 0.5$ with $\leq 0.5$, or $> 0.5$ with $\geq 0.5$: for if $\langle \neg\mathcal{A} \rangle_{\mathfrak{M}} = 0.5$, then $\langle \mathcal{A} \rangle_{\mathfrak{M}} = 0.5$, but on the corresponding classical interpretation $\mathfrak{M}_c$, only one of $\mathcal{A}$ and $\neg\mathcal{A}$ can be true.

31 Recall that the De Morgan laws hold in the algebra of degree functions, and note that if an algebra satisfies the conditions set out in Section 5, then the following generalised De Morgan laws also hold: $(\bigvee\{f_i\})' = \bigwedge\{f_i'\}$ and $(\bigwedge\{f_i\})' = \bigvee\{f_i'\}$.

32 For details, see for example (Keefe, 2000, pp. 175–176).

33 (a) An anonymous referee objected here that for soundness to be useful, it only needs to be the case that we can tell *to a pretty good approximation* whether it is satisfied in a particular case, and that this requirement is compatible with having *DF*'s in all their complexity figure in the definition of soundness. I *agree* that for soundness to be useful, it only needs to be the case that we can tell to a pretty good approximation whether it is satisfied in a particular case (this is exactly the situation according to my own view: supposing a sentence has a *DF* as its truth value, we can in general tell what this *DF* assigns to the empty sequence only *to a pretty good approximation*; we cannot in general know *exactly* which value it assigns), but I deny precisely that this is compatible with having *DF*'s in all their complexity figure in the definition of soundness: supposing a sentence has a *DF* as its truth value, in general we have *no idea at all* what this *DF* is like beyond the first couple of levels (what does it assign to sequences of length 100, or 1,000, or 1,000,000? – in general, no-one has the slightest idea).

(b) The reader may be wondering why I insisted that first approximations should not be given a privileged role when it comes to determining the ordering of *DF*'s (see footnote 24), and yet now I assert that they should be given a privileged role in the definition of validity and soundness. The two cases are fundamentally different and there is no reason to think they should be treated similarly: earlier we were concerned solely with relationships between *DF*'s; now we are also concerned with relationships between the sentences we use and their *DF*'s.

34 Interesting issues in philosophy of language arise here, but I cannot take them up in this paper. In case these issues have occurred to the reader, however, I should point out that what I have just said is quite compatible with the view that even before Cantor's work, the extensions of the terms 'has fewer elements', 'equinumerous', and so on, were completely fixed, even amongst infinite sets, and that Cantor did not *make* it true that there are the same number of even positive integers as positive integers, and would indeed have said something false had he denied this.

35 Again, this is not to say that we could not make an incorrect choice: the greater usefulness of one choice may simply be *evidence* of its correctness.

36 Tautologies are the sentences that have the tautology property on *every* interpretation.

37 If $\Gamma \models_v \mathcal{A}$ and $\Delta, \mathcal{A} \models_v \mathcal{B}$ then by Theorem 1, $\Gamma \models \mathcal{A}$ and $\Delta, \mathcal{A} \models \mathcal{B}$, whence $\Gamma, \Delta \models \mathcal{B}$ and hence by Theorem 1, $\Gamma, \Delta \models_v \mathcal{B}$.

38 If the parties to the discourse are epistemicists, then the statement is not odd at all. I am assuming however that the parties to the discourse are ordinary speakers, whose intuitions about vagueness are captured by the Closeness characterisation. For discussion of this point see (Smith, 2001, §2.1).

39 This would indicate that the real source of the intuition that the former is true to a high degree is understanding 'If Bob is bald, then Bill is bald' as meaning that if one were to stipulate a sharp boundary for 'bald', and it enclosed Bob, then it must enclose Bill also.

40 Although the Sorites paradox is traditionally presented in a form that involves conditionals, this is in no way essential.

41 If $\langle\mathcal{B}\rangle \geq 0.5$, then $\langle\mathcal{A} \rightarrow \mathcal{B}\rangle = \langle\neg\mathcal{A} \vee \mathcal{B}\rangle \geq 0.5$. If $\langle\mathcal{B}\rangle < 0.5$, then $\langle\mathcal{A}\rangle < 0.5$, so $\langle\neg\mathcal{A}\rangle > 0.5$, so $\langle\mathcal{A} \rightarrow \mathcal{B}\rangle = \langle\neg\mathcal{A} \vee \mathcal{B}\rangle > 0.5$.

42 See (Kleene, 1952, p. 334).

43 Here we see that we could not extend the original valuation scheme by analogy with Kleene's *strong* three-valued logic (Kleene, 1952, p. 334). Suppose $[\mathcal{A} \vee \mathcal{B}]_{\mathfrak{M}} = f$. Then on the strong extension, it could not be that $[\mathcal{A}]_{\mathfrak{M}} = *$ and $[\mathcal{B}]_{\mathfrak{M}} = *$, but *one* of them might have value $*$, and the other have value $f$. The $*$ value need not be preserved in $\mathfrak{M}'$; indeed it might be replaced by a value $g \neq f$ such that $f \vee g = g$, in which case $[\mathcal{A} \vee \mathcal{B}]_{\mathfrak{M}'} = g$, and monotonicity is lost.

44 See (Visser, 1989, p. 656).

45 In future I will leave it to the reader to sort out the different meanings of '$\leq$': this will not be hard, as in each context there will be only one thing that it sensibly can mean.

46 See (Visser, 1989, p. 657, Lemma 2.7(iii)).

47 The following three proofs are based on (Barker, 1998, pp. 17–18), who cites (Visser, 1989).

48 Note that in the semantics just presented, certain paradoxical sentences such as the following, which is the referent of the name '*a*'

$$\neg T_{\mathbf{T}} a$$

get the value $*$. Now intuitively, *a* says that it has a value other than **T**, and $*$ is a value other than **T**, so shouldn't *a* have the value **T**? To discuss this question, and all the further questions that the discussion would raise, would be to discuss the paradoxes of self-reference, and such a discussion is well beyond the scope of this paper. My attitude in the present section is that I simply do not care about paradoxical sentences such as the Liar. My aim was to show that we can make good on the intuition that when I say that Bob is bald to degree 0.7, this is just an approximation, and you might well say that my statement is true to degree 0.8 – and someone might say that your statement is true to degree 0.4, and so on. This intuition concerns the interrelation of non-pathological truth-talk and vague talk; what happens when we also mix in self-reference is another question altogether. If I were concerned with the latter question, I would need to consider other strategies for dealing with truth-talk, apart from Kripke's; for my purposes, however, there is no reason to look further than Kripke's construction: it may not provide the best account of the Liar paradox, but it does provide a good basis for the demonstration that ditto, approximate and definite truth predicates may consistently be introduced into the formal system presented in this paper, in such a way that they behave exactly as we want them to (in non-self-referential contexts – which is the sort of context we had in mind).

49 See (Williamson, 1994, §7.2) and (Williamson, 1992).

50 In fact Williamson is not exclusively concerned with many-valued semantics – his target is broader. As an advocate of many-valued semantics, however, this is how his argument strikes me.

51 For discussion of this issue see (Smith, 2001, §2.7) and (Smith, 2003), where I argue that the supervaluationists' argument from penumbral connection against the fuzzy view gets things precisely the wrong way around. Contra the supervaluationist, if (for example)

$x$ is a point on the rainbow midway between red and orange, ordinary speakers have exactly the same sort of hedging response to the claims '$x$ is red and $x$ is orange' and '$x$ is red or $x$ is orange' as they have to the claims '$x$ is red' and '$x$ is orange'.

52 That is, if they have a truth value (i.e. a value other than $*$) at all.

53 See (Smith, 2001, §3.7) for references to expressions of this intuition in the literature.

54 See (Smith, 2001, §§3.6.3, 3.8.2) and (Smith, 2003) for discussion of the issue of vague versus non-vague theories of vagueness.

55 I also assume that the binary predicate '$R$' is precise: if $a$ is not the immediate predecessor of $b$, then '$aRb$' is true to degree **F** (even if $a$ is, say, the immediate predecessor of the immediate predecessor of $b$).

56 See footnote 51.

57 See footnote 17.

58 The term 'correct interpretation' is due to (Islam, 1996). I prefer Islam's term to the term 'intended interpretation' because it does not give the impression that it must be the speaker's intentions that single out the privileged interpretation.

59 Fuzzy sets of type $n$ were introduced by Zadeh (1975a), and studied (in most cases only up to type 2) by (amongst others) Mizumoto and Tanaka (1976a, 1976b, 1981), Nieminen (1977), Yager (1980, 1984) and Hisdal (1981). Recently there has been renewed interest in type 2 fuzzy sets – in particular in their applications: see for example (John, 1998; Karnik et al., 1999) and (Liang and Mendel, 2000) (and references therein). In the philosophical literature, (Copeland, 1997) discusses fuzzy sets of higher type in relation to the question of the semantics of 'definitely' and 'indefinitely' operators.

60 This principle is introduced in (Zadeh, 1975a, pp. 236ff); it is also presented in (Zadeh, 1975b, pp. 416–420). It is discussed in most textbooks on fuzzy logic, for example (Dubois and Prade, 1980, pp. 36ff; Klir and Yuan, 1995, pp. 44–48) and (Nguyen and Walker, 2000, p. 30). I follow the formulation in (Dubois and Prade, 1980, p. 37).

61 This is the standard way of defining higher-type fuzzy operations, but in fact it is not the only possible way. Zadeh (1975a, 1975b) and Mizumoto and Tanaka (1976b) use the method just outlined; Hisdal (1981) presents an alternative method.

62 As far as I am aware, no-one in the literature has considered *sequences* of type $n$ fuzzy sets, let alone the restrictions on these sequences that would be needed to get something approaching my view.

63 The main sources are (Zadeh, 1975a, 1975b).

64 In fact I have written only of type 2 fuzzy *sets*, but I trust that the extension of the terminology to logics is obvious.

65 Here we see that Zadeh's terminology is somewhat unfortunate: he means by 'fuzzy subset of [0, 1]' a function from [0, 1] to [0, 1], and yet he withholds the term 'fuzzy' from [0, 1]-valued logic.

66 Recall that $\mathcal{T}$ is countable, while there are uncountably many fuzzy subsets of [0, 1].

67 Cf. Klir and Folger, as quoted by John in his article on type 2 fuzzy sets: "it may seem problematical, if not paradoxical, that a representation of fuzziness is made using membership grades that are themselves precise real numbers" (John, 1998, p. 563).

68 Note that Zadeh's idea of linguistic approximation of fuzzy subsets of [0, 1] by members of $\mathcal{T}$ is completely unrelated to the sort of approximation just mentioned.

69 Fine (1975, pp. 144–145) sketches a related construction within the context of three-valued logic.

## REFERENCES

Barker, J., 1998: The inconsistency theory of truth, Ph.D. thesis, Department of Philosophy, Princeton University.

Cantor, G., 1915: *Contributions to the Founding of the Theory of Transfinite Numbers*, Dover, New York. Translated by Philip E. B. Jourdain.

Copeland, B. J., 1997: Vague identity and fuzzy logic, *J. Philos.* **94**, 514–534.

Dubois, D. and Prade, H., 1980: *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York.

Fine, K., 1975: Vagueness, truth and logic, in (Keefe and Smith, 1997), pp. 119–150.

Hisdal, E., 1981: The IF THEN ELSE statement and interval-valued fuzzy sets of higher type, *Internat. J. Man-Machine Studies* **15**, 385–455.

Islam, A., 1996: The abstract and the concrete, Typescript.

John, R., 1998: Type 2 fuzzy sets: An appraisal of theory and applications, *Internat. J. Uncertainty, Fuzziness and Knowledge-Based Systems* **6**, 563–576.

Karnik, N. N., Mendel, J. M. and Liang, Q., 1999: Type-2 fuzzy logic systems, *IEEE Trans. on Fuzzy Systems* **7**, 643–658.

Keefe, R., 2000: *Theories of Vagueness*, Cambridge University Press, Cambridge.

Keefe, R. and Smith, P. (eds.), 1997: *Vagueness: A Reader*, MIT Press, Cambridge, MA.

Kleene, S. C., 1952: *Introduction to Metamathematics*, D. Van Nostrand, Princeton.

Klir, G. J. and Yuan, B., 1995: *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, Upper Saddle River, NJ.

Kripke, S., 1975: Outline of a theory of truth, *J. Philos.* **72**, 690–716.

Liang, Q. and Mendel, J. M., 2000: Interval type-2 fuzzy logic systems: Theory and design, *IEEE Trans. on Fuzzy Systems* **8**, 535–550.

Mizumoto, M. and Tanaka, K., 1976a: Fuzzy-fuzzy automata, *Kybernetes* **5**, 107–112.

Mizumoto, M. and Tanaka, K., 1976b: Some properties of fuzzy sets of type 2, *Inform. and Control* **31**, 312–340.

Mizumoto, M. and Tanaka, K., 1981: Fuzzy sets of type 2 under algebraic product and algebraic sum, *Fuzzy Sets and Systems* **5**, 277–290.

Nguyen, H. T. and Walker, E. A., 2000: *A First Course in Fuzzy Logic*, 2nd edn, Chapman & Hall/CRC, Boca Raton.

Nieminen, J., 1977: On the algebraic structure of fuzzy sets of type-2, *Kybernetika* **13**.

Rosen, G. and Smith, N. J. J., 2004: Worldly indeterminacy: A rough guide, *Australasian J. Philos.* **82**(1).

Smith, N. J. J., 2001: Vagueness, Ph.D. thesis, Department of Philosophy, Princeton University. UMI ProQuest Digital Dissertations no. AAT 3021982 (`http://wwwlib.umi.com/dissertations/fullcit/3021982`).

Smith, N. J. J., 2003: Vagueness and degrees of truth, forthcoming.

Visser, A., 1989: Semantics and the Liar paradox, in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, Vol. IV, D. Reidel Publishing Company, pp. 617–706.

Williamson, T., 1992: Vagueness and ignorance, in (Keefe and Smith, 1997), pp. 265–280.

Williamson, T., 1994: *Vagueness*, Routledge, London.

Wright, C., 1987: Further reflections on the Sorites paradox, *Philos. Topics* **15**, 227–290.

Yager, R. R., 1980: Fuzzy subsets of Type II in decisions, *J. Cybernet.* **10**, 137–159.

Yager, R. R., 1984: Fuzzy subsets with uncertain membership grades, *IEEE Trans. Systems Man Cybernet.* **14**, 271–275.

Zadeh, L. A., 1975a: The concept of a linguistic variable and its application to approximate reasoning – I, II, III, *Inform. Sci.* **8–9**, 199–249, 301–357, 43–80.
Zadeh, L. A., 1975b: Fuzzy logic and approximate reasoning, *Synthese* **30**, 407–428.

*Department of Philosophy,*
*and Centre for Logic, Language and Computation*
*Victoria University of Wellington*
*PO Box 600*
*Wellington*
*New Zealand*
*e-mail: nicholas.smith@vuw.ac.nz*