1

A Site for Sorites

Graham Priest

1. Introduction

The Sorites is a very hard paradox; I think it is harder than the paradoxes of self-reference. A measure of this is the fact that the paradoxes of self-reference may be solved by abandoning the law of non-contradiction. This has seemed to some a very drastic solution. But even this cannot help solve the Sorites: Sorites arguments lead not just to contradictions that can be isolated, but to wholesale contradiction. Using Sorites arguments one can prove nearly everything. For example, we can prove that you are scrambled egg as follows. Let b_0 be you, and suppose that there are n molecules in your body; let $b_0, b_1, \ldots, b_n$ be a sequence of objects each of which is obtained from its predecessor by replacing one molecule of you with a molecule of scrambled egg, so that b_n is all scrambled egg. Let β_i be the statement that you are b_i. Then clearly β_0, and for any i, $\beta_i \rightarrow \beta_{i+1}$. Hence by n applications of modus ponens β_n: you are scrambled egg.

I used to think that solving the Sorites paradox was an issue of logic; specifically, that one needed to give an account of the conditional that

Versions of this paper were given at a meeting of the Australasian Association of Philosophy held at the University of Canterbury, Christchurch, New Zealand, in July 2002, at the conference 'Liars and Heaps' held at the University of Connecticut, Oct. 2002, and at seminars at the University of St Andrews, Dec. 2002. I am grateful to the many members of the audiences on those occasions for many helpful thoughts and criticisms. The commentator on the second of these occasions was Achille Varzi, whose reply also occurs in this volume. In fairness to Achille, I have not modified the paper in response to his interesting comments. It is not easy to hit a moving target; further discussion can take place elsewhere.

showed why the argument fails, though we are ineluctably drawn to it. But, this, I now think, is false. The Sorites phenomenon has nothing, at root, to do with conditionals, validity, or any other apparatus of the logician. The phenomenon arises at a much more fundamental level, one prior to the engagement of any logical paraphernalia. It arises simply because we are forced to recognize the existence of cut-off points where both common sense and philosophical intuition scream that there are none. Thus, the only thing left for a solution to do is to explain why we find the existence of a cut-off so counter-intuitive. This is the site at which a solution to the Sorites needs to be sought.

The paper has three parts. The first explains why, I take it, we are stuck with cut-offs. The second formulates the beginnings of an explanation for this fact. The third discusses some issues that need to be addressed for this explanation to become a fully adequate solution to the Sorites paradox.

2. The Existence of Cut-Offs

2.1 The Forced March Sorites

The existence of a cut-off of some kind is forced upon us by a version of the Sorites sometimes called the 'forced march Sorites', which is, in fact, very close to the original version of the argument.[1] We can formulate this as follows. Consider the Sorites argument above. Let q_i be the question 'Is it the case that β_i?' If asked this question, there is some appropriate range of answers. What these are exactly does not matter. They might be 'yes', 'no', 'I don't know', 'yes, probably', 'er . . . ', or anything else. All that we need to assume is that an appropriate answer is justified by the objective state of affairs; specifically, by the nature of b_i.[2] Now, suppose I ask you the sequence of questions: $q_0, q_1, \ldots$ Given any question, there may be more than one appropriate answer. For example, you might say 'yes'; you might same 'same answer as last time' (having said 'yes' last time). All I insist is that

[1] The term 'forced march Sorites' was a coined (as far as I know) by Horgan (1994, sect. 4). The version I give here is slightly different from, and, it seems to me, tougher than, the version he gives there. His version is a metalinguistic identity Sorites. In due course, it will become clear how to deal with this. For the original formulation of the argument, see Williamson (1994, ch. 1) and Keefe and Smith (1997, ch. 2).

[2] The justification here is semantic, not epistemic. The answerer is personified simply to make the situation graphic.

once you answer in a certain way you stick to that until that answer is no longer appropriate. Suppose that in answer to the question q_0, you answer a. This may also be justified in answer to q_1, q_2, and so on. But there must come a first i where this is no longer the case, or it would be justified in answer to q_n, which it is not. Thus, for some i, b_{i-1} justifies this answer; b_i does not. The objective situation therefore changes between b_{i-1} and b_i in such a way.

And this, of course, is where intuition rebels. How can a single molecule make a difference? And by changing the example, we can make the difference between two successive states in a Sorites progression as small as one pleases. The difference may therefore fall below anything that is cognitively accessible to us; but vague predicates just don't seem to work like this.

Of course, how we should theorize this cut-off is another matter. Different theorists do this in different ways. A cut-off may be theorized as a change from truth to falsity; a change from truth to neither truth nor falsity, or to both truth and falsity; a change from being 100 per cent true to less than 100 per cent true; a change from maximal degree of assertibility to less than maximal degree; and so on. But never mind the details. What the forced march Sorites demonstrates is that any solution must face the existence of a cut-off. It cannot disappear it. All that is left for a solution to do is to theorize the nature of the cut-off, and to explain why we find its existence so counter-intuitive. This is the only form that a solution to the Sorites can take.

Let us consider a couple of replies. Here is one. The existence of a cut-off point seems odd because of the apparently arbitrary nature of its location. Suppose that the correct answer in the forced march Sorites changed at *every* question. The arbitrariness, and so the oddness, would then disappear. How could this be? One possibility is that to answer the question I simply *show* you the object at issue—which is changing from stage to stage. Another is that an answer is of the form 'It is true to degree r'—as in fuzzy logic—where r is a different real number every time.

The response to the first suggestion is fairly obvious. Let us not quibble about whether responding by showing is linguistic. If it is, the language is not ours. The problem into which the Sorites leads us is posed by the use of *our* language with its vague predicates, questions, and answers. We want a solution that applies to *that* language. One response to the second suggestion is similar. Even though, in this, the response to the question is by saying, not showing, a language with an uncountably infinite number of replies is not ours. But I think that there are greater problems with this response. However one conceptualizes degrees of truth, there are Sorites progressions where truth value does not change all the time. Thus, even if you were changed by replacing one molecule of your body with a molecule of scrambled egg, you would still be as you as you could be. You change more than that every morning after breakfast. Similarly, dying takes time, and so is a vague notion. But when your ashes are scattered to the four winds—and thereafter, if not before—you are as dead as dead can be. And if a correct answer to the relevant question does not change at every point, we face a counter-intuitive cut-off.

A second reply is to the effect that the answerer may "refuse to play the game". Of course, if they do this for subjective reasons, such as the desire to be obstreperous, this is beside the point. They might, however, do so for a principled reason, namely that the rules of the "game" are impossible to comply with. They lead the answerer, at some point, into a situation where they cannot conform. Now, it would certainly appear that it is possible to comply with the rules at the start: the first few answers present no problems. But then we may simply ask them to play the game as long as it is possible. If the answer changes before this, the point is made. If, however, they stop at some point before this, it must be because the situation is such as to require them both to give and not give the same answer as before. This was not the case at the question before, so the semantic situation has changed at this point. The only other possibility is for the answerer to say that the game in unplayable right at the start. But this can only be because there is no appropriate answer they can give even in the first case—and presumably, therefore, in all subsequent cases. This is not only implausible; it means that even in the most determinate case there is no answer that can be given. We are led to complete and unacceptable semantic nihilism.

2.2 Epistemicism and Contextualism

Of the solutions to the Sorites paradox currently on the market, very few address the counter-intuitiveness of the cut-off point explicitly. Perhaps the one that might be thought to do so most naturally is the epistemicism of Sorensen and Williamson.[3] Sorensen and Williamson subscribe to classical logic. In effect, then, they take all predicates to be semantically precise ones. In a sense, there are no vague predicates. In a soritical progression there is therefore a precise cut-off where the sentences turn from true to false or vice versa.

[3] See Sorensen (1988, esp. 189–216) and Williamson (1994, esp. chs. 7, 8).

The distinctive feature of epistemicism is an attempt to explain why we find the existence of the cut-off point counter-intuitive in terms of features of our knowledge. We do not know where it is—indeed, there are reasons why we cannot know where it is: the location of the cut-off is in principle unknowable. This is why we find its existence counter-intuitive. It should be noted that, though Sorensen and Williamson deploy epistemicism in defence of classical semantics, it could be deployed equally in defence of *any* of the standard semantics for vagueness. As I have noted, they all entail the existence of cut-off points, and all, therefore, are in need of an explanation of why this is counter-intuitive. Epistemic considerations of the kind in question can be invoked.

Williamson's explanation of the in-principle unknowability of the cut-off point can be put in various different ways, but at its heart it depends on the fact that knowledge supervenes on an evidential basis. If two situations are effectively the same in the evidence that they provide (are relevantly similar) then I cannot know something about one but not about the other. In particular, then, if I know something about one such situation, that thing must be true in the other situation too—or, being false, I would not know it. Williams calls this the 'margin of error principle'. In particular, suppose that $\beta_0, \beta_1, \ldots, \beta_n$ is a soritical sequence of statements. Suppose that β_i and all prior members are true, that β_{i+1} and all subsequent members are false, and that one knows where the cut-off is, i.e. one knows that β_i is true and that β_{i+1} is not. Since I know that β_i is true, by the margin of error principle, β_{i+1} must be true too—which, *ex hypothesi*, it is not.

Though epistemicism is in the right ball-park for an explanation, it does not stand up well to the cold light of inspection. A major worry is that the very phenomenon that explains why we cannot know where the cut-off point is undercuts its very existence. The meanings of vague predicates are not determined by some omniscient being in some logically perfect way. Vague predicates are part of *our* language. As a result, their meanings must answer in the last instance to the use that *we* make of them. It is therefore difficult to see how there could be a semantic cut-off at a point that is *in principle* cognitively inaccessible to us. Of course, cut-off points for crisp predicates may be inaccessible to us too in a certain sense. Thus, if an electron accelerates uniformly from rest to some velocity, there must be a precise instant at which it has half that velocity. Even if we know the terminal velocity in question, we may not be able to determine that instant, due to the limitations of our measuring instruments, etc. But the in-principle unknowability of the cut-off points for vague predicates is quite different from this. In the case of the electron we know exactly what it is about the world that determines where the cut-off point resides, and why *that* particular fact settles the matter. In the case of the vague predicate, we have neither of these things. To suppose that such exists would appear to be a form of semantic mysticism.[4]

Worse, and crucially in the present context, it is doubtful, in the end, that epistemicism can explain why we find the existence of cut-off points so counter-intuitive. There are many things that we cannot know and whose existence we do not find puzzling in the same way at all. For example, there is a well-known model of the physical cosmos according to which the universe goes through alternating periods of expansion and contraction. In particular, the singularity at the big bang was just the end of the last period of contraction and the beginning of the current period of expansion. Suppose this is right. Then there must be many facts about what happened in the phase of the universe prior to the big bang—for example, whether there was sentient life. Yet all information about this period has been wiped out for us—lost in the epistemic black hole that is the big bang. Yet we do not find the existence of determinate facts before the big bang counter-intuitive. Indeed, we seem to have no problem imagining there to be such things, though they are and ever will be cognitively inaccessible to us. To explain why we cannot know the existence of something does not, therefore, explain why we find its existence counter-intuitive.

Another account of vagueness that might be thought to lend itself to explaining why we find the existence of precise cut-off points counter-intuitive is contextualism, of the kind proposed by Graff (2000). For present purposes, the pertinent features of the view are as follows. First, vague predicates are contextually dependent. This is plausible: 'tall' for a basketball player and 'tall' for a pygmy certainly have different extensions. Second: what is psychologically salient is part of the context. This seems plausible too. When we point in a general direction, and say 'that', we expect the hearer to take the referent of the demonstrative to be whatever is salient in that direction. Third, and most importantly here, is what Graff calls the *similarity constraint*: if, given a fixed context, two salient objects are relevantly similar with respect to a predicate, then it applies to both or neither. This, too, is not implausible: it is a localized version of the natural idea that vague predicates are tolerant with respect to small changes. Given these ideas, we might try to explain why we find the existence of a cut-off point counter-intuitive as

[4] As Crispin Wright puts it his detailed critique of epistemicism (1995). See also Horgan (1994, sect. 5).

follows. Whenever we look to find the relevant cut-off at a certain point, this thrusts those objects on either side of the point into salience. Thus, because of the similarity constraint, the cut-off point relative to that context is not there. In other words, wherever we look for the cut-off point it is not there. So we come to believe that there isn't one.

Again, one may have worries about this sort of account of vagueness quite generally. Can't I have a very short Sorites, say with a handful of colour strips, where it is clear that the endpoints have different colours, and yet where every strip in the context is salient? After all, I may not be able to get a hundred strips into mental focus at the same time, but I would certainly seem to be able to get five or six. But setting this aside, it is not clear that contextualism really does explain why we find the existence of a cut-off counter-intuitive. Suppose that we look to find the cut-off at some point or other. The similarity constraint explains why it is not there. With respect to that context it must therefore be elsewhere—outside the area of salience. But the thought that there is a sharp cut-off point somewhere else is still as puzzling as before. The tolerance of vague predicates seems, after all, to be a general phenomenon, not simply localized to the area of salience. What could make *that* the cut-off point is therefore just as puzzling as before.

At least as applied in the most obvious ways, then, neither epistemicism nor contextualism provides an explanation as to why we find the existence of a cut-off counter-intuitive. What could provide an explanation? Conceivably there could be many putative explanations. Maybe epistemicism or contextualism can be deployed in some other way. Maybe an appropriate explanation could be purely psychological: some deep psychological mechanism produces the illusion of cut-off freedom. And if we are lucky enough to come up with a number of viable explanations, we will have to determine which is the best. But at the moment, this is a non-issue; for presently we have none. In the next part of the paper I want to sketch one possible explanation. This is neither epistemic nor purely psychological, but logical.

3. An Explanation

3.1 Vague Identity

The explanation is parasitic on fuzzy logic—though fuzzy logic is not, in itself, a solution to the problem, as I have already noted. The logic is harnessed to allow a certain explanation of the counter-intuitiveness of a precise cut-off point. For the sake of definiteness, let us suppose that the logic is determined by the continuum-valued semantics of Łukasiewicz—though other fuzzy logics could also be employed. The details of the logic are well enough known for me not to have to repeat them here.[5] I remind you of just one fact. For something to be acceptable, it does not have to have unit truth-value. A value high enough will do. In practice, the context will determine some value $0 < \varepsilon < 1$, such that a value greater than or equal to ε is sufficient to make a sentence acceptable. The inferences that preserve acceptability whatever value of ε is chosen are precisely those whose conclusions are always at least as true as (the conjunction of) their premises.[6]

Now, let us start by recalling that there are Sorites arguments that depend not on modus ponens but on the substitutivity of identicals.[7] Consider a soritical sequence of colour patches such that each is phenomenologically indistinguishable from its immediate neighbours, which begins with a pure shade of red and ends with a pure shade of blue. For the sake of definiteness, let us suppose that there are 100 such patches. Let c_i be the colour of the ith patch. Clearly, $c_1 = c_1$, and for all $1 \leq i < 100$, $c_i = c_{i+1}$. By a sequence of substitutions $c_1 = c_{100}$.

This argument shows that identity itself must be a fuzzy predicate; identity statements must therefore have degrees of truth. An appropriate semantics for identity is as follows.[8] The domain of quantification is furnished with a normalized metric. This is a map, d, from pairs of objects to real numbers satisfying the conditions:

$$0 \leq d(x, y) \leq 1$$
$$d(x, x) = 0$$
$$d(x, y) = d(y, x)$$
$$d(x, y) \leq d(x, z) + d(z, y).$$

An identity statement, $a = b$, has degree of truth $1 - d(\mathbf{a}, \mathbf{b})$, where $\mathbf{a}$ is the denotation of a, etc. The members of the domain, as furnished with the appropriate notion of identity, can be thought of as fuzzy objects, objects whose identity comes by degrees.

[5] See e.g. Priest (2001, ch. 11).

[6] See Priest (2001: 11.4.10).

[7] See Priest (1991).

[8] See Priest (1998). Note that that paper reverses the usual conventions concerning 1 and 0 as degrees of truth.

It is not difficult to show that the transitivity of identity is not (globally) valid in these semantics. Indeed, a model of the colour Sorites is provided by the interpretation where:

$$d(\mathbf{c}_i, \mathbf{c}_j) = \frac{|i-j|}{100}.$$

The truth-values of $c_1 = c_2$ and $c_2 = c_3$ are 0.99; the truth-value of $c_1 = c_3$ is 0.98. Hence the inference $c_1 = c_2,\ c_2 = c_3 \vdash c_1 = c_3$ is invalid.

Although the transitivity of identity is invalid, there is a weaker notion of validity that it satisfies: local validity. Loosely, an inference is locally valid if its conclusion will be acceptable provided that its premises are *true enough*. A crucial property of locally valid inferences is that a chain of locally valid inferences is itself locally valid, though the degree of truth required by the premises becomes higher and higher the longer the chain is. Consequently, locally valid inference can be used a few times to make conclusions acceptable, but prolonged use is liable to end in something unacceptable. (Details of how to make these ideas formally precise can be found in Priest 1998.) Not only is the transitivity of identity invalid but locally valid; so is modus ponens.

3.2 Fuzzy Truth-Values

With this background, we can now turn to the required explanation. The idea is, very simply, to take semantic values themselves to be fuzzy objects. Thus, the relevant domain of quantification is the set of real numbers between 0 and 1 (thought of as fuzzy real numbers), and we take the metric on these to be the standard distance metric $|x - y|$. To explain how things work, it will be easiest to take a simple example. Suppose that $\alpha_0, \ldots, \alpha_9$ is a soritical sequence of sentences whose semantic values are as shown in the following table. I write the truth-value of α_i as $\tau(\alpha_i)$. The second row tabulates the value of the sentence $\tau(\alpha_0) = \tau(\alpha_i)$.

α_i	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
$\tau(\alpha_i)$	1	1	1	0.8	0.6	0.4	0.2	0	0	0
$\tau(\alpha_0) = \tau(\alpha_i)$	1	1	1	0.8	0.6	0.4	0.2	0	0	0

Let us suppose that the cut-off for acceptability, ε, is 0.7.

As the forced march Sorites requires, there is a first sentence where $\tau(\alpha_i)$ is distinct from $\tau(\alpha_0)$. In fact, this is α_4, since 4 is the first i for which

$\tau(\alpha_0) = \tau(\alpha_i)$ is not acceptable (less than 0.7). So far, no surprises. But now, and crucially, why do we find the existence of such a cut-off counter-intuitive? As is easy to check, the truth-values of $\tau(\alpha_3) = \tau(\alpha_4)$ and $\tau(\alpha_4) = \tau(\alpha_5)$ are both 0.8, which is greater than 0.7. Thus the value of α_4 is identical with that of each of its neighbours! This is why we find the existence of the cut-off counter-intuitive! In fact, as is easy to check, for all $0 \leq i \leq 9$, the value of $\tau(\alpha_i) = \tau(\alpha_{i+1})$ is ≥ 0.8 (the bound being obtained at several places). Hence the value of the sentence $\forall i\ \tau(\alpha_i) = \tau(\alpha_{i+1})$ is 0.8 too. Hence, every sentence has the same truth-values as its neighbours. Note that we cannot use this fact to show that truth-values are identical all the way down the Sorites, since the transitivity of identity fails.[9]

The precise numbers employed in the preceding example are, of course, simply illustrative. The general point that they illustrate is that the truth-value of the sentences must change eventually as we go down the sequence, despite the fact that it never changes as we go from each sentence to its neighbour. Unlike the epistemic solution to the Sorites, this does not fly in the face of common sense. It is exactly what common sense seems to tell us! What this solution gives us is an account of how the trick is turned. Not all sequences make a cut off-point counter-intuitive, of course. Thus, suppose we consider three stages of a person's life: age less than 10, age 10–20, and age over 20. Let us ask at which stage they change from being a child to not being a child; the answer is in the second stage, and there is nothing counter-intuitive about this. The idea of the solution just adumbrated is that when the existence of a cut-off point *is* counter-intuitive, the distribution of truth-values and the relevant level of acceptability conspire to make it so in the way indicated.[10]

[9] A little computation shows that the value of the sentence $\exists i(\tau(\alpha_i) = \tau(\alpha_0) \wedge \tau(\alpha_{i+1}) \neq \tau(\alpha_0))$ is less than 0.7. It might therefore be suggested that the model does not verify the existence of a cut-off point. The sense in which there is a cut-off point is that there is an i such that $\tau(\alpha_i) = \tau(\alpha_0)$ is acceptable and $\tau(\alpha_{i+1}) = \tau(\alpha_0)$ is not. To express this in the language we would need to extend it, as we could do, to express the thought that a sentence is acceptable.

[10] Let me comment on one other feature of the present proposal. It is frequently objected to fuzzy logic that the supposition that a sentence has a real-valued truth-value—with all its infinite precision—is itself highly counter-intuitive. The theory of fuzzy truth-values can be seen as addressing this problem too. The truth-value of a sentence is a fuzzy real. If, suppose, it is 0.7, it may equally be 0.65 and 0.75.

4. Further Issues

4.1 Degrees of Truth and Local Validity

So much for the idea. In the final sections of the paper I will reflect on issues that need to be faced if this solution to the Sorites is to be articulated into something that is fully adequate.

Suppose that one has a language with vague predicates, including identity, L_1. We may, in a metalanguage, give a theory, T, which provides an account of a continuum-valued semantics for the language (including the non-transitive semantics for identity). While the underlying logic of T would normally be left at an informal level, a natural enough assumption would be that this is classical logic. All this is orthodox and routine enough.

In one sense, the preceding construction can be seen as a particular example of this. The language in question, L_1, has names for the sentences of some language, L_0 (which might or might not be the same is L_1), and for real numbers; it contains function symbols, including $\mathbf{\tau}$ and those naturally interpreted as expressing arithmetic functions; finally, it contains an identity predicate, and possibly various others. We may give a semantics for this language in the way just described. In a sense, this is all that is presupposed in the preceding considerations.

However, as will be clear, the particular language, L_1, in question is not any old language. It is itself a language containing metalinguistic notions. Hence, one might reasonably expect that it be not only the language whose semantics is in question, but the vehicle for the metalinguistic theory about L_0 itself. If we think of it in this way, we can certainly no longer take the logic of the metatheory to be classical logic: it must be a fuzzy logic of some kind. What ramifications does this have?

The matter is far from straightforward. Part of the problem is that there are at least two distinct purposes for which metalinguistic reasoning may be engaged, though these are often not clearly distinguished. One is in an investigation of truth; the other is in an investigation of validity. Let us take these two matters in turn.

One matter for which metalinguistic reasoning is employed is to articulate a notion of truth, and establish certain things as true or otherwise. In the present case, it is not truth *simpliciter* that is at issue, but degree of truth. Thus, we suppose that L_1 contains axiom schemata such as:

(Neg) $\tau(\langle\neg\alpha\rangle) = 1 - \tau(\langle\alpha\rangle)$

where, now that we are being more precise about the syntax, $\langle.\rangle$ is an appropriate name-forming device. Given that we also have other axioms that specify the degrees of truth of atomic sentences (or allow us to infer them from more fundamental axioms concerning the denotations of their parts), we may infer the degrees of truth of more complex sentences.

Now the crucial question: what underlying logic is required for this project? A full answer to this question can be provided only by working out the project in detail, for which this is not the place. But the amount of logic required would seem to be surprisingly little. Suppose, for example, that we have established that, for some particular α, $\tau(\langle\alpha\rangle) = 0.5$. The computation of the value of $\neg\alpha$ goes as follows:

$\tau(\langle\alpha\rangle) = 0.5$	Already established
$\tau(\langle\neg\alpha\rangle) = 1 - \tau(\langle\alpha\rangle)$	Neg
$\tau(\langle\neg\alpha\rangle) = 1 - 0.5$	SI
$1 - 0.5 = 0.5$	Arithmetic theorem
$\tau(\langle\neg\alpha\rangle) = 0.5$	TI

Here, SI and TI are, respectively, the substitution of identicals and the transitivity of identity.

Though the argument deploys a somewhat minimal amount of logic, the most notable thing in the present context is that the argument uses inferential moves concerning identity that are not valid, notably TI. This is the first issue that needs to be faced. The problem is not, in fact, as desperate as it may at first appear. Even though TI is not valid, it is locally valid. This means that, though we cannot use it over "long distances", it is perfectly all right to employ short chains of inference involving it. (It should be noted that though SI—of which TI is a special case—is not valid, an appropriate form is also locally valid.[11])

This raises two questions. The first, and most obvious, is how short 'short' is. The answer to this depends on factors such as how true the initial axioms are, what degree of truth is acceptable, and so on. While these questions demand to be addressed in detail, it can be said that if things are set up right, the short may be very long—maybe longer than any chain of inference that anyone can construct in practice. The invalidity may not, therefore, be a practical problem.

[11] See Priest (1998, sect. 6).

Eventually, though—wherever eventually is—the acceptability of the arguments may give out. The second question is what to make of this. The obvious answer is that our (meta)theoretical machinery will cease to be sufficient to determine degrees of truth for certain sentences. Perhaps this is something that can simply be faced with equanimity. Maybe there just is no fact of the matter concerning the degrees of truth of such sentences. If these sentences are *really* long—say, longer than anything constructible in practice, and so transcending anything humanly meaningful—this is perhaps not implausible.

4.2 Kicking Away the Ladder

The second purpose for which metalinguistic reasoning may be employed is in an investigation, not of truth, but of validity. A metatheory is used to formulate a semantic notion of validity for languages of a certain kind, and to investigate what inferences, couched in those languages, are and are not valid. In addition, we may hope to isolate a set of proof procedures and show them to be sound, and maybe also complete, with respect to the semantic notion.

What of the means of inference employed in these investigations themselves? Suppose that the language of the metatheory is itself a language of the kind in question. Then if the metatheoretic reasoning is to be understood as giving an account of, or justification for, those inferences that are acceptable for languages of this kind, the inferences employed should not be ones that are demonstrably invalid. More generally, one might hope, one should be able to demonstrate the soundness of the class of inferences employed.

This is the situation we are in if we take the language L_1 as a fuzzy language in which an account of fuzzy validity is to be given. How do things stand with this? A definition of validity is easy enough to formulate. Sticking to the one-premise case, for simplicity, we can define the validity of an inference with premise α and conclusion β, $V(\langle\alpha\rangle\langle\beta\rangle)$, as follows:

$$V(\langle\alpha\rangle\langle\beta\rangle) \leftrightarrow \forall\upsilon(\upsilon(\langle\alpha\rangle) \leq \upsilon(\langle\beta\rangle)).$$

But what inferences should one be allowed to use to reason about this notion of validity? It is clear, for a start, that it is no longer to be expected that this is classical logic. The whole point of applying fuzzy logic to the Sorites is that inferences like modus ponens and TI are not valid. A better guess is that we take it to be fuzzy logic itself. Such a logic, however, is too weak. Establishing that various inferences are valid, let alone establishing general soundness (and completeness) results, is going to require inferences such as TI, SI, and modus ponens. Merely consider, for example, the natural argument to show that the inference from $\alpha \wedge \beta$ to α is valid:

$\upsilon(\langle\alpha \wedge \beta\rangle) = Min(\upsilon(\langle\alpha\rangle), \upsilon(\langle\beta\rangle))$	Truth condition
$Min(\upsilon(\langle\alpha\rangle), \upsilon(\langle\beta\rangle) \leq \upsilon(\langle\alpha\rangle))$	Arithmetic theorem
$\upsilon(\langle\alpha \wedge \beta\rangle) \leq \upsilon(\langle\alpha\rangle)$	SI
$\forall\upsilon(\upsilon(\langle\alpha \wedge \beta\rangle) \leq \upsilon(\langle\alpha\rangle))$	Universal generalization

It is therefore clear that the arguments employed must include locally valid inferences as well.

This raises the same kind of issue that we met in the last section. In particular, there is a question of the extent to which such inferences are acceptable in this context. It may also mean that certain inferences are neither demonstrably valid nor demonstrably invalid. Maybe, then, there is no fact of the matter here.

But a new question is also posed. Is reasoning of this kind sufficient to show the validity (including local validity) of the very reasoning in question? That is, can the soundness of this reasoning be demonstrated employing the reasoning itself? Moreover, but less crucially, can it also establish it to be appropriately complete. No doubt it can using classical logic. Can the classical ladder be kicked away? The question cannot be answered without detailed investigation, but the following comments are in order.

As it is not uncommon for bright students to observe, demonstrating the validity of a principle of inference tends to employ the very principle in question (together with some fairly basic logical apparatus, such as modus ponens, etc.—which we have, in this case, though it is only locally valid). One would therefore expect that most sensible logics would be able to establish the validity of any inference it holds to be valid. This much, at least, ought to be possible in the case at hand. And this is enough to ensure that the logic is self-coherent.

Establishing a formal soundness theorem requires somewhat more by way of reasoning; for example, it requires that one can define the set of axioms and reason appropriately about its members. This requires a certain amount of set theory—at the very least the ability to define certain sets by recursion. And the corresponding completeness proof requires even more—that one can show the existence of various counter-models and demonstrate their properties. This does not necessarily require classical logic. Recall that there are intuitionistically valid soundness and completeness proofs for intuitionistic propositional logic.[12] But constructing such proofs is not at all a routine matter. Witness the problems around the production of an intuitionistically valid completeness proof for intuitionistic predicate logic.[13]

5. Conclusion

The issues that I have raised in the last part of the paper all need to be addressed to make the solution to the Sorites that I have suggested fully articulated. Some of these are distinctly non-trivial. For this reason, I do not claim that the solution I have suggested is right. It does, however, seem to me to be one that is both plausible and worthy of further investigation.

REFERENCES

DUMMETT, M. (1977), *Elements of Intuitionism* (Oxford: Oxford University Press).

GRAFF, D. (2000), 'Shifting Sands: An Interest-Relative Theory of Vagueness', *Philosophical Topics*, 28: 45–81.

HORGAN, T. (1994), 'Robust Vagueness and the Forced-March Sorites', *Philosophical Perspectives*, 8: 159–88.

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

PRIEST, G. (1991), 'Sorites and Identity', *Logique et Analyse*, 135–6: 293–6.

—— (1998), 'Fuzzy Identity and Local Validity', *The Monist*, 81: 331–42.

—— (2001), *Introduction to Non-Classical Logic* (Cambridge: Cambridge University Press).

SORENSEN, R. (1988), *Blindspots* (Oxford: Oxford University Press).

WILLIAMSON, T. (1994), *Vagueness* (London: Routledge).

WRIGHT, C. (1995), 'The Epistemic Conception of Vagueness', *Southern Journal of Philosophy*, 33: 133–59.

[12] See e.g. Dummett (1977, esp. 214).

[13] See Dummett (1977, sect. 5.6, esp. p. 259).