Nicholas K. Jones, Birkbeck College   Nick.K.Jones@gmail.com

On Supervaluations, Meaning and Consequence.

Abstract.

The problem of vagueness is motivated by consideration of a sorites series, and the supervaluationist solution is presented. It is argued that supervaluationism can be regarded as an application of the model-theoretic conception of consequence and should therefore be accepted by all. The limitations of this conception are then explored. In particular, if the theory is to provide a constitutive account of vagueness, an account of vagueness as underspecification is required.

1 – Vagueness, sorites and borderline cases.

Vagueness is problematic because it appears to give rise to *sorites arguments*. To motivate these arguments, we begin by constructing a *sorites series* for 'red'.

Consider the white tiles α_1, α_2, …, α_n and n empty paint tins. The first tin is filled with n-1 measures of scarlet paint, the second with n-2 measures of scarlet paint and 1 measure of yellow paint, the third with n-3 measures of scarlet and 2 measures of yellow paint, …, and the nth with n-1 measures of yellow paint only. The tins are stirred thoroughly, then α_1 is dipped in the first, α_2 in the second, …, and α_n in the nth. Once the paint has dried α_1 is clearly red and α_n is clearly not red[1]. The tiles are arranged in a series with α_1 at the origin, α_n at the terminus and with each α_i the successor of α_{i-1}. When n is large enough (500, say), successive tiles are indiscriminable with respect to colour[2]. Our concern is exclusively with series where n is large enough. Though the vagueness of English constitutes our primary

---

[1] For simplicity, assume that the object-language is impoverished relative to English, insofar as it contains no word for a shade intermediate between red and yellow. What English speakers would call 'orange', the speakers of the object-language lack a word for. Thus the orange tiles are neither clearly red nor non-red. Nothing hangs on this oversimplification: the example may be modified to involve any two shades between which English recognises no intermediate shade.

[2] Whether x is indiscriminable from y is relative to the individual attempting to make the discrimination, background conditions under which the discrimination is made, mode of presentation of the candidates, and respect in which the discrimination is made. Consider the concept of discriminability used throughout to be implicitly relativised to a typical human in normal background conditions, where individuals are presented visually and discriminated between on the basis of their colour. Note that indiscriminability is non-transitive.

interest, in the interests of presentation and clarity, we translate into the formal language of standard first-order predicate logic with identity, interpreted on a domain of individuals α_1, …, α_n. The predicate 'is red' is translated as F and names for the tiles are as in English. Premiss 1 in the following sorites argument is clearly true[3].

1. $F\alpha_1$.

2. $\forall x_i\ (Fx_i \rightarrow Fx_{i+1})$.

⋮ ─────────────────

$n+1$ $\therefore F\alpha_n$.

Line $n+1$ is clearly false and follows from premisses 1 and 2 by n-1 applications of *modus ponens*. Since premiss 1 is clearly true, it follows by *reductio* that premiss 2 is false. Premiss 2 is the major, or *sorites*, premiss.

(¬2) $\exists x_i\ (Fx_i\ \&\ \neg Fx_{i+1})$.

(¬2) is the negation of the sorites premiss and follows classically from its falsity. If we take the sorites premiss to be false, then we should endorse the truth of (¬2). But, intuitively, (¬2) also appears false, so we should not endorse it either. So we both should and should not endorse (¬2). Some way must be found to accommodate rejection of both the sorites premiss and (¬2).

(¬2) appears false because the series contains borderline cases: those α_i, α_{i+1}, …, α_j where $1 < i \leq j < n$, that are neither clearly red nor clearly non-red (clearly yellow). Predications of 'red' to these individuals are *borderline predications*. (¬2) asserts that there is some α_i to which 'is red' applies succeeded by some α_{i+1} to which 'is red' does not apply. Given the principle (T), bivalence and a suitable abstraction principle for forming predicates from sentences, it follows that 'is not red' (the complement of 'is red') applies to α_i, by substituting 'α_i is red' in place of 'S'.

(T) 'S' is true iff S.[4]

[3] Contextual variation in semantic values is ignored throughout. It is assumed that sorites arguments can be assessed relative to a single context.

[4] This is a variant of an argument due to Timothy Williamson (1994: 187-98). As Williamson remarks, bivalence, truth and falsity apply only to utterances that say something. 'S' could be neither true nor false (α_i

So the pair $\{\alpha_i, \alpha_{i+1}\}$ marks a cut-off point between the red and non-red individuals that lies between successors. Borderline cases appear to rule out there being any such pair. They create a borderline region in the series that separates the last F from the first non-F, thus apparently making (¬2) false. But (¬2) follows classically from premisses 1 and 2 with the negation of (the clearly false) line $n+1$ in the sorites argument above[5].

The problem of vagueness is to find a response to sorites arguments. We might reject classical logic or bivalence, but it would be methodologically improper to do so without having first explored the alternatives. Endorsing (¬2) is problematic because of the existence of borderline cases. An investigation of borderline cases is therefore required. In English, the borderline cases of 'is red' are those individuals that are neither clearly red nor clearly not red. We translate 'clearly' into our formal language as the sentential operator 'Δ'[6]. The borderline cases (of ϕ) are those individuals x that satisfy schema (B).

(B) $\qquad \neg\Delta\phi x \ \& \ \neg\Delta\neg\phi x.$

Vagueness requires rejection of (2) and therefore endorsement of (¬2). Endorsement of (¬2) is problematic because of borderline cases. Borderline cases are characterised using 'Δ'. So an account of 'Δ' is a compulsory component of an adequate theory of vagueness. This theory should deliver (at least) two things.

First, we require inference rules for 'Δ', so that we can reason about and in the presence of vagueness. These rules should allow us to explain why (¬2) is unproblematic, despite appearances to the contrary. Second, we should like to know what vagueness *is*. Is it a semantic, metaphysical, epistemological, psychological or pragmatic phenomenon? Once one of these is settled on, an account of the features of that domain that give rise to vagueness should be provided.

---

satisfy neither 'red' nor 'not red') if it does not say anything. Throughout, it is assumed that all utterances dealt with have propositional content.
[5] If line $n+1$ were not clearly false, sorites arguments would show that all vague predicates are trivial – apply to either all or no individuals, given the assumption that they apply (or not) to some individual. Since they are not trivial, line $n+1$ is false.
[6] The English 'definitely', 'clearly' and 'determinately' are used interchangeably throughout and hence all formalised as 'Δ'.

2 – Supervaluations.

Currently, the most popular treatment of vagueness employs supervaluations. This is a formal technique developed by Bas van Fraasen and first applied to vagueness by Kit Fine and David Lewis (van Fraasen (1966), Fine (1975), Lewis (1970a), Williamson (1994): 142-164 provides an influential recent presentation and Keefe (2000) a recent defence).

The initial idea is that a vague language admits of multiple equally good ways of being made precise; multiple *admissible precisifications*. Precisifications assign (classical) semantic values, of the appropriate kind, to each expression in the language. A sentence is definitely true iff it is true on all admissible precisifications, and definitely false iff it is false on all admissible precisifications. Sentences that are true on some, but false on other, admissible precisifications are neither definitely true nor definitely false. These are the borderline sentences. The borderline cases are those individuals that satisfy some borderline sentence. Each precisification is classical, so classical logic is preserved.

It follows from each precisification's assignment of classical semantic values (and *F*'s being non-trivial) that (¬2) is true on each and therefore definitely true. But since no instance of (¬2) is true on each admissible precisification (we could draw the borderline for red in many different equally good places) no instance is definitely true. Our reluctance to endorse (¬2) is explained as a weak intuitive grasp on scope distinctions and the inference rules for Δ. We make an invalid inference from the absence a definite instance of (¬2) to its being definitely false. Really, says the supervaluationist, borderline status does not imply falsity and the truth of (¬2) does not require a definitely true instance. A sorites series as described above is consistent with the definite truth of (¬2) (represented as (D) below), for that requires only a true instance on each admissible precisification, not some one instance that is true on each (as represented by (R) below). (D) is consistent with the negation of (R), though speakers of English mistakenly reason as if it were not. So the definite truth of (¬2) is consistent with its lacking any definite instance.

(D) $\qquad \Delta \exists x_i\, (Fx_i\ \&\ \neg Fx_{i+1})$.

(R) $\qquad \exists x_i\, \Delta (Fx_i\ \&\ \neg Fx_{i+1})$.

Slightly more formally, we can see this as providing an interpretation of the possible world semantics for modal logic.

Begin with the notion of a *specification space* (analogous to a domain of worlds). This is a set of points. Each point provides a classical interpretation of the object-language; assigns elements of the domain to singular terms and sets of ordered n-tuples of elements to n-place predicates. The interpretation of the logical apparatus is constant across points. Each point defines a model of the object-language. A relation of *admissibility* holds between points and provides a partial ordering of them. This is a primitive of the space; its extension is fixed by constraints imposed by the language the space is provided for. The first constraint is that of *penumbral connections*: non-logical connections between predicates (any red thing is coloured, any individual taller than a tall individual is a tall individual etc). The second constraint is that of *truth-preservation*: the truth-value of any definite sentence is preserved under precisification (if the Incredible Hulk is definitely tall, then any point at which 'the Hulk is tall' is false is not admissible). The definite sentences at a point are those true at every point it admits. 'Δ' is thus treated as formally akin to the '□' of modal logic. As Williamson (1994: 156-162) has pointed out, admissibility should be non-transitive, thus invalidating both the characteristic S4 and S5 axioms, and making space for higher-order vagueness (borderline cases to the borderline cases)[7].

3 – A case for supervaluations.

There is a sense in which the method of supervaluations should be accepted by all parties. Recall that one adequacy criterion on a theory of vagueness is that it must provide inference rules for 'Δ'. These rules tell us what are, and are not, the logical consequences of a sentence prefixed by 'Δ', and when it may be introduced. The notion of logical consequence can be assimilated to the notion of logical necessity, and explicated in a framework akin to that described above.

Iff the inference from S to T is valid, T is true at every model of the language at which S is also true (Tarski 1936, Etchemendy 1999: 24). To find the logical consequences of S, we find every model in which S is true and then see what else holds true in each of them. Hold fixed the truth-value of S and vary the interpretation of the rest of the language (consistently with the truth of S). Points in the specification space and models in the account of consequence play the same role.

---

[7] Williamson's criticisms of the supervaluationist treatment of higher-order vagueness will not be addressed here.

This framework is formally identical to that of supervaluationism. The truth value of sentences prefixed with 'Δ' is held constant (true), and the interpretation of the language varied in accordance with that. The logical consequences of 'ΔS' are those sentences true under every interpretation under which 'ΔS' is true. In examining the inference rules for 'Δ', hold fixed the truth of the sentence within its scope.

By way of example, consider '$\Delta \exists x \, (Fx)$'. In investigating the logical consequences of this, we hold fixed the truth of '$\exists x \, (Fx)$' and vary its interpretation. The truth of '$\exists x \, (Fx)$' is compatible with variation in the assignment of individual to 'x'. Each assignment results in a different true instance of 'Fx', so no sentence of the form '$F\alpha$' need be true in every model in which '$\exists x \, (Fx)$' is. So the inference from '$\Delta \exists x \, (Fx)$' to '$\Delta F\alpha$' is invalid.

Contrast this with '$\exists x \, \Delta(Fx)$', where 'Δ' lies within the scope of '$\exists x$'. We must instantiate to the true instance of this quantification in order to obtain a sentence, '$F\alpha$', prefixed by 'Δ'. Holding the truth of this fixed, while varying its interpretation obviously retains the truth of '$F\alpha$'. So the inference from '$\exists x \, \Delta(Fx)$' to '$F\alpha$' is valid.

We see that the supervaluationist account of 'Δ' and the logician's understanding of consequence coincide. Both hold fixed the truth-value of a sentence and then vary the interpretations of the rest of the language in accord with it. The supervaluationist constraint of truth-preservation is built into this technique, and penumbral connections can be accommodated by treating object-language expressions of them as definite. The supervaluationist account of the inference rules for 'Δ' is thus available to all.

There is a problem. How are we to use this technique to discover all the inference rules for 'Δ' when we do not already know some? An extreme case would be if 'ΔS' entails some sentence T but, intuitively, the vocabulary in appears T unrelated to that in S. In this case there would be appear to be no constraints imposed by the truth of S at a model on the truth of T at that model, and the supervaluationist technique would deliver the wrong results. Unless we already know something about the connection between S and T, the inference from S to T need not be ruled valid by the supervaluationist. The reply is that supervaluationism reveals minimal inference rules for 'Δ'. We know that the inference from 'ΔS' to 'S' is valid. This is undeniable. Given this, we can find other rules for 'Δ'. There may be others that slip through the net, but these should be seen as further rules. Adding premisses cannot make a valid inference invalid. Supervaluationism provides a minimal set of inference rules that

other accounts should agree to also (unless they show that supervaluationism makes unwarranted presuppositions about the language; e.g. that no vague language should receive a wholly classical interpretation)[8]. Given the minimal assumption that the move from 'ΔS' to 'S' is valid, other minimal rules can be found also.

4 – Meaning.

We have seen that the formal structure of supervaluations is the same as that in which logical consequence is explicated. To this extent, supervaluationism should not be denied as a theory of vagueness. This does however raise the question of how we should view the theory. What is the relevance of precisifications to the semantics of natural language?

If each expression of the language is actually assigned some semantic value, then that value determines the truth-value of the sentence at all models that respect its meaning. There can be no variation in semantic value without variation in meaning. If the semantic value of 'F' is varied across precisifications, then it does not receive its intended interpretation at many of them, in which case the supervaluationist is guilty of changing the subject. Admissible ways of making the language precise do not change its meaning. But a change in meaning is required for a change in semantic value. So precisifications are irrelevant to the meaning and hence inference rules of 'Δ'.

Ernest LePore and Jerry Fodor (1996) claim that the semantic value of an expression is a matter of conceptual necessity. The suitability of semantic values, model-theoretically conceived, supervenes upon the meaning of an expression. There can be no variation in semantic value without variation in the expression itself. In which case, a theory that treats linguistic items as receiving different semantic values at different models (as does supervaluationism) is guilty of changing the topic. It is concerned with a range of non-English languages, or predicates other than 'is red', not with the vague predicate with which we began. If an expression receives a truth-value upon semantic evaluation, then it receives that value upon all evaluations, and if it does not, then it does not upon all evaluations. Any

---

[8] One particularly striking consequence is that if, following Fine and Keefe, we identify the truth-predicate with 'Δ', then the logical connectives are not truth-functional when they connect borderline sentences. When a and b are duplicates with respect to F, Fa and Fb receive the same truth-value. When both are exactly midway between the clear and clear non-cases, $\neg Fa$ and $\neg Fb$ receive the same truth-value as Fa and Fb. But $\neg Fa$ & Fb is clearly false, while Fa & Fb is a borderline case. It is this that motivates the constraint of *penumbral connections*.

evaluation that claims to assign it a different truth-value is one on which it is conflated with some alternative expression. Because meaning is a matter of conceptual necessity, we cannot conceive of precisifications even as possible meanings of an expression – they vary semantic value and are therefore not possible meanings of the original expression.

These arguments assume a view of the meanings of vague expressions that is not uncontroversial. It is common among supervaluationists to conceive of vagueness as a special kind of ambiguity (Lewis (1993) in particular). Vague terms have their meaning underspecified in the sense that the meaning-determining facts are not sufficient to fix on any one semantic value. Borderline predications are those predications true under some, but not all, ways in which the linguistic community might make the unmade decision as to what the meaning of the vague term in question is to be. On this view, vagueness is not a feature of the meaning of an expression, but a deficiency. So long as precisifications do not vary those aspects of semantic value that have been settled, and vary only the unsettled cases, they are not guilty of changing the topic. Since the undecided cases are not settled by the semantic value of the linguistic items in question, deciding them does not count as *varying* the expression, and hence not as changing the subject.

This approach accords well with John Etchemendy's (1999) modification of Tarski's account of consequence. Tarski held that only the interpretation of the logical vocabulary should be held constant across models, but lacked any systematic account of how to determine which the logical constants are. Etchemendy proposes that in order to overcome this, we are allowed to hold fixed the interpretation of whichever class of expressions we are interested in. Logical consequence is the result of holding only the logical vocabulary fixed, but this can be generalised to other forms of consequence by holding other expressions fixed also. Since it is the definite truths we are interested in, we need hold fixed only those aspects of the semantic values of expressions that are definite – settled by the meaning-determining facts. This is just what the supervaluationist does. Thus, if we are prepared to allow that meanings are underspecified, supervaluationism and the modified account of consequence coincide.

Prima facie it did not appear that supervaluationism is committed to a view of meaning as semantic underspecification. What we have seen is that there is a direct argument for supervaluationism, but it succeeds only given this view. If, following Fine and Keefe, we construe supervaluations as part of the semantic mechanism by which truth-values are

determined (and hence identify 'Δ' with the truth-predicate) this argument will not be available, for on that view meanings are fully specified. Either meanings are incompletely specified and we have a direct argument for supervaluationism, or they are fully specified and the argument here is for a theory that is guilty of changing the topic.

5 – The theory as a tool.

There is a response available. Thus far, we have seen supervaluationism as providing a realist theory of how truth-conditions for sentences prefixed with 'Δ' are determined. Precisifications and the admissibility relation between them were genuine components of the mechanism by which sentences with vague expressions are semantically evaluated. We might not do so.

Instead, we could choose to take supervaluationism as providing an instrumental tool for reasoning about, and in the presence of, vagueness. As we saw above, it provides a structure in which to determine the inference rules of 'Δ'. This is just the structure used to determine the consequences of any expression. The structure need not be taken as a constitutive account of what vagueness *is*. That matter may be left entirely open. Maybe vagueness is a semantic phenomenon, maybe metaphysical, maybe epistemic, but whatever it is, it provides 'Δ' with the inference rules the supervaluationist proposes (and thereby makes available the supervaluationist solution to the sorites). Our argument for supervaluationism can be made consistent with a picture of meanings as fully specified if we are willing to forego an account of the underlying nature of vagueness and settle with a logic for 'Δ'.

This leaves us with two conceptions of supervaluationism. Both provide us with inference rules for 'Δ' and a solution to the sorites. By regarding every aspect of the theory as representational, we obtain a constitutive account of vagueness at the cost of being forced to regard vague expressions as underspecified and acknowledge an unparsimonious profligacy of meanings/languages. Alternatively, we can regard supervaluationism as providing only a formal structure. In this sense it is trivially correct. But we should not regard precisifications as having anything to do with the meaning of English, only its logic, and hence lack a constitutive account of vagueness. This conception of the theory as a tool should not place strong constraints on the inference rules for 'Δ' presented by other, realistically construed, theories of vagueness. It is consistent with regarding supervaluationism as a tool for

determining inference rules only that it makes some incorrect predictions in its initial form. Upon discovery of the incorrect predictions, modification of the structure to avoid them is required.

An interesting example is provided by sentences of the form:

(V) $\exists X \, \exists x \, (\neg \Delta Xx \, \& \, \neg \Delta \neg Xx)$.

Supervaluationism makes (V) trivially false (Williamson 2003). Variables refer rigidly across precisifications/models to their semantic values; once their values are fixed, so is the truth-value of the relevant instance of 'Xx' at all models; it cannot have that value other than definitely. This result can be avoided by allowing variables to refer non-rigidly across models, but at the cost of losing the solution to sorites arguments[9]. (V) asserts that there are vague states of affairs in Reality: some object neither determinately has nor lacks some property. This is ontic vagueness. Supervaluationism is inconsistent with it.

Degree-theory, by contrast, makes (V) true whenever a sentence lacks a determinate truth-value. The reason is that consideration of only one semantic value, is required in order to secure the borderline status of a predication. If a predication is borderline, that is because it states a borderline fact. All vagueness is ontic/metaphysical. By construing degree-theory realistically, we can accommodate the supervaluationist structure as a formal tool. The tool must be modified to allow variables to refer non-rigidly, and a degree-theoretic response to the sorites adopted. Thus we retain the supervaluational framework by regarding it non-realistically and adopting an alternative realistic account. In this way supervaluationism can be made consistent with metaphysical vagueness.

---

[9] The reason is complex and I will not go into it here. See Williamson (2003) for full details.

References.

Etchemendy, J. (1999), 'Reflections on consequence'. Unpublished MS available from http://www-csli.stanford.edu/hp/Reflections.pdf

Fine, Kit (1975), "Vagueness, truth and logic" in *Synthese* 30: 265-300. Reprinted in Keefe & Smith (1997) 119-150.

Keefe, R. & Smith, B. (1997) (editors), *Vagueness: a reader*. MIT Press.

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

LePore, E. & Fodor, J. (1996), 'What cannot be evaluated cannot be evaluated, and it cannot be supervalued either' in *Journal of Philosophy* 93: 516-36.

Lewis, D. (1970a), 'General Semantics' in *Synthese* 22: 18-67, reprinted in Lewis, D. (1983a).

Lewis, D. (1983a), *Philosophical Papers: Volume 1*. New York, Oxford: OUP.

Lewis, D. (1993), 'Many but almost one' in Campbell, K., Bacon, J. & Reinhardt, L. (1993), reprinted in Lewis (1999): 164-182.

Lewis, D. (1999), *Papers in Metaphysics and Epistemology*. Cambridge, New York, Melbourne: Cambridge University Press.

Loux, M. & Zimmerman, D. (2003) (eds.), *The Oxford Handbook of Metaphysics*. Oxford: OUP.

McGee V. & McLaughlin, B. (1994), 'Distinctions without a difference' in *Southern journal of philosophy* 33: 203-245.

Tarski, A. (1936), 'O pojciu wynikania logicznego" in *Prezglad Filozoˉczny* 39: 97-112. Translated as 'On the Concept of Logical Consequence' in Tarski (1956): 409-420.

Tarski, A. (1956), *Logic, Semantics, Metamathematics*. OUP.

van Fraasen, B. C. (1966), 'Singular terms, truth-value gaps and free logic' in *Journal of Philosophy* 63: 481-95.

Williamson, T. (1994), *Vagueness*. London, New York: Routledge.

Williamson, T. (2003), 'Vagueness in Reality' in Loux. J & Zimmerman, D. (2003): 690-715).