Liars and Heaps

New Essays on Paradox

EDITED BY
JC Beall

CLARENDON PRESS · OXFORD

Raffman, D. (1996), 'Vagueness and Context-Relativity', *Philosophical Studies*, 81: 175–92.

Rawls, J. (1955), 'Two Concepts of Rules', *Philosophical Review*, 64: 3–32.

Rooth, M. (1985), 'Association with Focus', Ph.D. diss., University of Massachusetts at Amherst.

——(1992), 'A Theory of Focus Interpretation', *Natural Language Semantics*, 1: 75–116.

Soames, S. (1989), 'Presupposition', in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, vol. iv (Dordrecht: Kluwer).

——(1999), *Understanding Truth* (Oxford: Oxford University Press).

Stalnaker, R. C. (1978), 'Assertion', in P. Cole (ed.), *Pragmatics*, vol. ix of *Syntax and Semantics* (New York: Academic Press); repr. in Stalnaker (1999).

——(1999), *Context and Content* (Oxford: Oxford University Press).

Strawson, P. F. (1950), 'On Referring', *Mind*, 59; repr. in Strawson (1971).

——(1952), *Introduction to Logical Theory* (London: Methuen).

——(1954), 'A Reply to Mr. Sellars', *Philosophical Review*, 63: 216–31.

——(1971), *Logico-Linguistic Papers* (London: Methuen).

Tappenden, J. (1993), 'The Liar and Sorites Paradoxes: Toward a Unified Treatment', *Journal of Philosophy*, 90: 551–77.

Urquhart, A. (1986), 'Many-Valued Logic', in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, vol. ii. (Dordrecht: Kluwer).

van Deemter, K. (1996), 'The Sorities Fallacy and the Context-Dependence of Vague Predicates', in M. Kanazawa, C. J. Piñón, and H. de Swart (eds.), *Quantifiers, Deduction, and Context* (Stanford, Calif.: CSLI Publications).

von Fintel, K. (forthcoming), 'Would you Believe It? The King of France is Back!', in A. Bezuidenhout and M. Reimer (eds.), *Descriptions and Beyond* (Oxford: Oxford University Press).

van der Sandt, R. A. (1992), 'Presupposition Projection as Anaphora Resolution', *Journal of Semantics*, 9: 333–77.

Wedgwood, R. (1997), 'Non-cognitivism, Truth, and Logic', *Philosophical Studies*, 86: 73–91.

Wiggins, D. (1980), 'What would be a Substantial Theory of Truth?', in Z. van Straaten (ed.), *Philosophical Subjects: Essays Presented to P. F. Strawson* (Oxford: Oxford University Press).

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

——(1996), 'Knowing and Asserting', *Philosophical Review*, 104: 489–523.

Wright, C. (1992), *Truth and Objectivity* (Cambridge, Mass.: Harvard University Press).

9

Gap Principles, Penumbral Consequence, and Infinitely Higher-Order Vagueness

Delia Graff

Philosophers disagree about whether vagueness requires us to admit truth-value gaps, about whether there is a gap between the objects of which a given vague predicate is true and those of which it is false on an appropriately constructed Sorites series for the predicate—a series involving small increments of change in a relevant respect between adjacent elements, but a large increment of change in that respect between the endpoints. There appears, however, to be widespread agreement that there is *some* sense in which vague predicates are gappy which may be expressed neutrally by saying that on any appropriately constructed Sorites series for a given vague predicate there will be a gap between the objects of which the predicate is *definitely* true and those of which it is *definitely* false. Taking as primitive the operator 'it is definitely the case that', abbreviated as 'D', we may stipulate that a predicate F is definitely true (or definitely false) of an object just in case '$DF(a)$', where a is a name for the object, is true (or false) *simpliciter*.[1] This yields the following conditional formulation of a 'gap principle':

I am grateful to Alexis Burgess, Michael Fara, Gilbert Harman, and Harold Hodes for helpful discussion of the topics in this paper, and also to audiences at the University of Connecticut, Oxford University, and the University of Rochester where this material was presented.

[1] Throughout, single quotes are used for 'scare' quotation, direct quotation, name-forming quotation, and also, instead of corner quotes, for 'quasi-quotation'.

$$(D\Phi(x) \wedge D\neg\Phi(y)) \rightarrow \neg R(x, y).$$

Here 'Φ' is to be replaced with a vague predicate, while 'R' is to stand for a *Sorites relation* for that predicate: a relation that can be used to construct a Sorites series for the predicate—such as the relation of *being just one millimetre shorter than* for the predicate 'is tall'. Disagreements about the sense in which it is correct to say that vague predicates are gappy can then be recast as disagreements about how to understand the definitely operator. One might give it, for example, a pragmatic construal such as 'it would not be misleading to assert that'; or an epistemic construal such as 'it is known that' or 'it is knowable that'; or a semantic construal such as 'it is true that'.

Those who think that the gappiness of vague predicates amounts to a gap between truth and falsity will also accept a formulation of the gap principle as an argument schema without a definitely operator. Since they believe that a verifier and a falsifier for a given vague predicate cannot be adjacent elements of a Sorites series for that predicate, they will accept, for example, that it is not possible for it to be true that x is tall and that y isn't unless it's true that x is more than one millimetre taller than y; that from x's being tall, together with y's not being tall, it *follows that* x is not just one millimetre taller than y. Schematically:

$$\frac{\Phi(x) \;\wedge\; \neg\Phi(y)}{\therefore \neg R(x, y)}.$$

In what follows I argue that acceptance of gap principles creates problems for those who maintain that the gappiness of vague predicates amounts to a gap between truth and falsity. In particular, I argue first that 'higher-order' gap principles, in their conditional formulation, lead to contradiction when the definitely operator is given a semantic construal; and second, that supervaluationists who accept gap principles in their formulation as arguments must concede that the sense in which they endorse classical logic is more qualified than is typically advertised.

1. The Paradox of Higher-Order Gap Principles

Some say that a predicate is vague just in case it has, or could have, borderline cases, where what it is to be a borderline case of a vague predicate such as 'tall' (for example) may be expressed using the definitely operator: to be borderline tall is to be neither definitely tall nor definitely not tall.[2] I prefer to say that a predicate is vague just in case it has fuzzy boundaries of application on an appropriately constructed Sorites series for the predicate. The metaphor of fuzzy boundaries can be understood in at least two ways, which are themselves in turn metaphorically expressed. On one understanding, a predicate has fuzzy boundaries of application just in case it has a *grey area* of applicability. On this understanding, to have a fuzzy boundary of application just is to have borderline cases. On another understanding of the metaphor of fuzziness, a predicate has fuzzy boundaries of application along a Sorites series when there is an apparently *seamless transition* along the series from cases where the predicate applies to cases where it doesn't apply. On this understanding, fuzziness is associated with susceptibility to paradoxical Sorites reasoning. I regard fuzziness understood as seamless transition as equally well described using Crispin Wright's (1975) notion of *tolerance*—the applicability of vague predicates seems to us tolerant of small differences or changes yet not always tolerant of large ones, which is paradoxical since small changes accumulate. I prefer to think of fuzziness *qua* seamless transition in terms of Mark Sainsbury's (1991*a*) notion of *boundarylessness*: although a vague predicate clearly does apply to some initial segment of its Sorites series but clearly doesn't apply to some final segment, the transition from the former segment to the latter seems boundaryless; the transition seems to occur along the series while seeming not to occur at any point in the series.

Crucially, the two conceptions of fuzziness come apart. There could seem to be a grey area of applicability on a Sorites series without an appearance of boundarylessness, since a grey area may itself appear sharply bounded. Conversely, there could in principle be an appearance of boundarylessness without there seeming to be any grey area at all. The metaphor of fuzziness may obscure this latter point, since it seems that a fuzzy boundary must occupy an extended region, and hence that to lie within this region is to be a borderline case. To that extent, the metaphor of fuzziness is not entirely apt.

The positing of a gap between a vague predicate's definite cases and its definite non-cases often seems, in theories of vagueness, to serve two functions, corresponding to the two conceptions of fuzziness just outlined. Consider the case of the predicate 'tall (for a man)'. On the one hand, there are men who seem to fall in the grey area—when asked whether they are tall, we feel that some kind of hedged answer would be most

[2] Although I do not think that gradable adjectives such as 'tall' have predicate-type semantic values, I regard it as harmless in this context to call them *predicates*.

appropriate; and moreover it seems that no further information would make us feel more confident in answering simply 'yes' or 'no'. These men falling in the grey area are classified as not definitely tall but not definitely not tall either. On the other hand, when confronted (in our imagination) with a Sorites series for the predicate, we find ourselves unable to locate any point marking the transition from the tall to the not tall. The proposed reason for this inability is that between the men who definitely are tall and those who definitely are not are some men who are neither. Thus a theory of definiteness on this picture is required simultaneously to be both a theory of what it is to fall in the grey area, i.e. cause us to hedge, and a theory of our inability to locate a boundary.

But we are no more able to locate a point marking a transition from the definitely tall to the *not* definitely tall than we are able to locate one marking a transition from the tall to the not tall. The push to accept higher-order vagueness comes from requiring that the very same explanation be given for this inability as was given for the first: just as we cannot locate a boundary dividing the tall from the not tall because there are men who do not fall definitely into either category between the men who fall definitely into one or into the other, we cannot locate a boundary dividing the definitely tall from the not definitely tall because there are men who do not fall definitely into either of *these* categories between the men who fall definitely into one or into the other. So although there is no obvious phenomenon of second-order hedging—anyone indecisive about whether to be indecisive about whether a given man is tall just is being indecisive about whether the man is tall; the claim that a given man is tallish-*ish* is not clearly even sensible—a theory of definiteness is now on this picture required to accommodate there being not only first-order borderline cases, men who are neither definitely tall nor definitely not tall, but also 'second-order borderline cases', men who are neither definitely definitely tall nor definitely not definitely tall.

The problem does not stop there, for we also cannot locate a boundary between the definitely definitely tall and the not definitely definitely tall (assuming, for the sake of argument, that we can understand what it would be to fall into these categories). So in addition to our first-order gap principle (in conditional formulation) for 'tall', we also have many higher-order gap principles.

$(DT(x) \wedge D\neg T(y)) \rightarrow \neg R(x,y)$ — Gap Principle for '$T(x)$'
$(DDT(x) \wedge D\neg DT(y)) \rightarrow \neg R(x,y)$ — Gap Principle for '$DT(x)$'
$(DDDT(x) \wedge D\neg DDT(y)) \rightarrow \neg R(x,y)$ — Gap Principle for '$DDT(x)$'
$\vdots$ $\vdots$

I will assume that we have classical equivalences governing the interaction of the conditional with conjunction and negation, so that the following, for example, are equivalent: $(\varphi \wedge \psi) \rightarrow \neg\theta$ and $\theta \rightarrow (\psi \rightarrow \neg\varphi)$.[3] By 'equivalent', I mean *must have the same truth value*, where for the purposes of this definition, a lack of truth value is stipulated to be a truth value. The equivalences in question are accepted by standard gap theorists, such as supervaluationists and also those who adopt the Kleene strong three-valued truth tables. Now, using prime symbol notation to denote the successor of a term in a Sorites series (following Crispin Wright), we may represent our gap principles schematically as follows, extended to cover any number of iterations of the definitely operator, as indicated by superscripts.

Gap Principle for $D^n\Phi(x)$: $DD^n\Phi(x) \rightarrow \neg D\neg D^n\Phi(x')$
Equivalent formulation: $D\neg D^n\Phi(x') \rightarrow \neg DD^n\Phi(x)$.

Some principles governing the logic of the definitely operator are uncontroversial, while others are appropriate only to particular construals of it. Some relatively uncontroversial principles are the following:

(T) $\vdash D\varphi \rightarrow \varphi$
(K) $\vdash D(\varphi \rightarrow \psi) \rightarrow (D\varphi \rightarrow D\psi)$
(RN) if $\vdash \varphi$ then $\vdash D\varphi$.

The semantic construal of the definitely operator, according to which it may be understood as akin to an operator such as 'it is true that', leads in addition to a principle to which epistemicists emphatically object: a rule of D-introduction. In its strong form the rule allows one to infer $D\varphi$ from some premises if one can infer φ from those same premises.

(D-intro) if $\Gamma \vdash \varphi$ then $\Gamma \vdash D\varphi$.

It is natural for a truth-value gap theorist to accept D-introduction, and in particular (since $\varphi \vdash \varphi$) to regard the inference from φ to $D\varphi$ as valid, given her construal of definiteness as *truth*. For it seems impossible for a sentence S to be true while another sentence—'it is true that S'—that says (in effect) that

[3] I'll freely use symbols, such as $\neg$ or D, as names of themselves; and also use concatenation itself, rather than the concatenation symbol $\frown$, to stand for the concatenation function so that, for example, $D\varphi$ = 'D'$\frown\varphi$ (using name-forming quotes) = '$D\varphi$' (using quasi-quotes) for any formula φ.

it's true is not true. Standard supervaluationist semantics and conceptions of validity such as that developed by Kit Fine (1975) and endorsed more recently by Rosanna Keefe (2000)[4] do indeed yield that *D*-introduction is validity-preserving.[5]

Crispin Wright (1987, 1992) has argued that acceptance of a second-order gap principle leads to paradox. Specifically, he has argued that given *D*-introduction one can deduce from a second-order gap principle—for 'definitely tall'—the following Sorites sentence: for all *x*, if the Sorites-successor of *x* definitely *isn't* definitely tall, then *x* definitely isn't definitely tall as well.[6] This is paradoxical: a three-foot tall man definitely *isn't* definitely tall, so appeal to the deduced Sorites sentence yields that the preceding man on the series, who is just one millimetre taller, also definitely isn't definitely tall. Repeated appeal to the Sorites sentence yields that the first man on the series, who, we may presume, is 2.5 metres tall, also definitely isn't definitely tall. But this is false. Any 2.5-metre tall man definitely *is* definitely tall.

With some minor variations, Wright's deduction is as follows (line numbers in square brackets track premises):

[1]	(1)	$\forall x(DDT(x) \to \neg D\neg DT(x'))$	Premise
[1]	(2)	$DDT(x) \to \neg D\neg DT(x')$	(1) $\forall$-elim
[3]	(3)	$D\neg DT(x')$	Premise (for conditional proof)
[4]	(4)	$DT(x)$	Premise (for reductio)
[4]	(5)	$DDT(x)$	(4) *D*-intro
[1,4]	(6)	$\neg D\neg DT(x')$	(2,5) $\to$-elim
[1,3,4]	(7)	$D\neg DT(x') \wedge \neg D\neg DT(x')$	(3,6) $\wedge$-intro
[1,3]	(8)	$\neg DT(x)$	(7)[4] $\neg$-intro
[1,3]	(9)	$D\neg DT(x)$	(8) *D*-intro
[1]	(10)	$D\neg DT(x') \to D\neg DT(x)$	(9)[3] $\to$-intro
[1]	(11)	$\forall x(D\neg DT(x') \to D\neg DT(x))$	(10) $\forall$-intro

Dorothy Edgington (1993) and Richard Heck (1993) respond by claiming that one cannot apply *D*-introduction within sub-proofs. As Heck formulates the response, *D*-introduction is validity-preserving, but in its presence rules such as conditional-introduction and negation-introduction that involve

[4] See also Heck (1993) and Williamson (1994, ch. 5).

[5] Vann McGee and Brian McLaughlin are supervaluationists who do not accept *D*-introduction because they think the principle conflicts with higher-order vagueness, but for reasons different from the ones brought out here (McGee and McLaughlin 1998, forthcoming).

[6] Wright formulates the deduction so that it appeals only to a weakened form of *D*-introduction: if $D\Gamma \vdash \varphi$ then $D\Gamma \vdash D\varphi$, where $D\Gamma = \{D\gamma : \gamma \in \Gamma\}$.

discharge of premises cease to be validity-preserving without restriction, with the result that premises may not be discharged via these rules once *D*-introduction has been applied to them or to something deduced from them.

I now want to argue, revisiting this decade-old debate,[7] that higher-order gap principles lead, given *D*-introduction, to contradiction. I will construct a new argument, based on the same idea as Wright's, that is not subject to Edgington and Heck's criticism.

We have a Sorites series of *m* objects (call the first '1', the second '2', and so on), each *R*-related to its successor in the series, beginning with an object we can truly describe as tall and ending with an object we can truly describe as not tall. Then the sequence of sentences below begins with a true sentence, while each subsequent sentence follows from the one immediately preceding it either by *D*-introduction or by appeal to a gap principle (second conditional formulation) and modus ponens. No discharge of premises is involved.

$\neg T(m)$	
$D\neg T(m)$	*D*-intro
$\neg DT(m-1)$	Gap principle for $T(x)$
$D\neg DT(m-1)$	*D*-intro
$\neg D^2T(m-2)$	Gap principle for $DT(x)$
$D\neg D^2T(m-2)$	*D*-intro
$\neg D^3T(m-3)$	Gap principle for $D^2T(x)$
$\vdots$	
$\neg D^{m-1}T(1)$	Gap principle for $D^{m-2}T(x)$

A further argument beginning with $T(1)$ yields $D^{m-1}T(1)$, after $m-1$ applications of *D*-introduction. Contradiction.

My proof that the truth of higher-order gap principles makes it impossible (given *D*-introduction) to get from an object that isn't tall, in any finite number of *R*-steps, to one that is tall exploits a rather obvious fact: it is not possible for any relation *R* to densely order only finitely many objects. We can make the point more vivid. Grasp the first member of a length-*m* Sorites series for 'tall' in your left hand; grasp the last member in your right hand. To illustrate that there's no 'sharp' boundary between the tall and the not-tall, you want to move your right hand leftward to grasp a different object that is a borderline case of the predicate 'tall' that's true of the object in your left hand but false of the object in your right hand.[8] After one move leftward of your

[7] See also Sainsbury (1991*b*).

[8] I equate falsity with truth of the negation.

right hand you still have the object in your left hand that is tall, hence definitely tall, and a new object in your right hand that's a borderline case of 'tall', hence *not* definitely tall. Now to illustrate that there is no sharp boundary between the definitely tall and the not definitely tall, you want to move your right hand leftward again, to grasp an object that's a borderline case of the predicate 'definitely tall' that's true of the object in your left hand but false of the object in your right hand. Each time you do this, you find you have an object in your left hand of which a predicate of the form '$D^nT(x)$' is true, and an object in your right hand of which that predicate is false. The collection of $m-1$ gap principles appealed to in my argument entails that you can do this at least $m-1$ times. But you cannot do this as many as $m-1$ times; there were only $m-2$ objects between your hands at the start.[9]

We should get clear about what exactly has and has not been shown. It has not been shown that there are higher-order predicates of the form '$D^nT(x)$' for which there is no true gap principle. Let me explain. Compare two Sorites series for the predicate 'tall', each beginning with a man whose height is 2.5 metres (definitely tall) and ending with a man whose height is 1.5 metres (definitely not tall). One series consists of 101 men, each one centimetre taller than the next; while the other series consists of 1,001, each one millimetre taller than next. All we have shown is that the truth of gap principles involving the successor relation from the first series must give out before we get to 100 iterations of the definitely operator, while the truth of gap principles involving the successor relation from the second series must give out before we get to 1,001 iterations of the definitely operator.[10] So it remains an open possibility that for any predicate of the form '$D^nT(x)$', involving any number of iterations of the definitely operator, there is *some* (perhaps very) finely discriminating relation involved in a true gap principle for that predicate.

That that possibility remains open, however, has no bearing on the point I wish to make. The point I wish to make is that given a *particular* finite-length Sorites series for the predicate 'tall', involving a particular Sorites relation for that predicate, we are no more able to locate a point on that series marking the transition from the objects of which 'definitely definitely... definitely tall' is true to those of which it is false than we are able to locate one marking the transition from the objects of which 'tall' is true to those of which it is false. The truth of higher- and higher-order gap principles for the particular series—involving its particular successor relation—cannot be what explains our inability to locate such points, since (given *D*-introduction) not all such principles are true.

We correspondingly have also not shown that there is no such thing as infinitely higher-order vagueness. I take it that on the picture we're working with a predicate is first-order vague if, on some Sorites series for it, there is a gap between the things of which it is definitely true and the things of which it is definitely false. I take it also that a predicate is infinitely higher-order vague if it is vague to some order but not to any greatest order. We have not yet said, however, what it is for a predicate to be vague to an order greater than one. I presume that the first-order vagueness of 'definitely tall' would suffice for the second-order vagueness of 'tall', and in general that the first-order vagueness of 'definitely^n tall' would suffice for the $(n+1)^{\text{th}}$-order vagueness of 'tall'. This does not amount to a definition, however,[11] since there may be other higher-order predicates, including 'definitely^n not-tall' and 'borderline^n tall', the first-order vagueness of which would also suffice for the $(n+1)^{\text{th}}$-order vagueness of 'tall'. But we won't need a general definition for the purposes of the discussion here; it will be enough to have identified the sufficient conditions for n^{th}-order vagueness already mentioned.

A gap principle for 'definitely^n tall' is (the universal closure of): $DD^nT(x) \rightarrow \neg D\neg D^nT(x')$. If for each such predicate there is *some* Sorites series—involving perhaps only very slight differences in height between adjacent elements—that furnishes us with a successor relation that renders the corresponding gap principle true, then, given the preceding considerations, that suffices for 'tall' to be infinitely higher-order vague. For if a gap principle for 'definitely^n tall' is true, relative to some given Sorites series, then nowhere on that series will we find an object which 'definitely^n tall' is definitely true succeeded by one of which it is definitely false. So if for every predicate of the form 'definitely^n tall' there is some such Sorites series, 'tall' is infinitely higher-order vague.

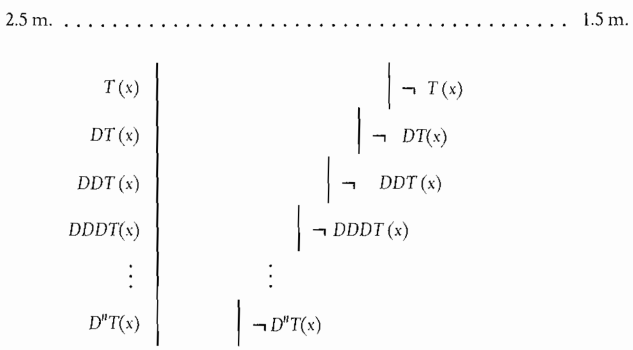
One might think, in light of the foregoing discussion, that at the very least we have shown that if 'tall' is infinitely higher-order vague, this cannot be

[9] A related challenge can be posed for Timothy Williamson's (1994) view that gaps in knowledge rather than gaps in truth value explain our inability to locate boundaries for vague predicates. See Gómez-Torrente (1997) for presentation, and Williamson (1997) for a reply. See also Gómez-Torrente (2002) and Graff (2002) for further developments of the challenge, and Williamson (2002) for a reply to those.

[10] This has been emphasized to me by Hartry Field, Stewart Shapiro, and Timothy Williamson.

[11] See Williamson (1999) and the introduction to Graff and Williamson (2002, p. xxi) for a general definition of n^{th}-order vagueness.

exhibited on any one finite-length Sorites series. Yet we have not shown even that much. What we have shown, using reasoning the typical truth-value gap theorist accepts, is that given a particular Sorites series for 'tall', not every higher-order gap principle formulated with the successor relation from that series is *true*. We have not, however, shown that a truth-value gap theorist must accept that any such principle is *false*. It is only when a gap principle for 'definitely^n tall' is *false* that we are assured of finding, on the Sorites series in question, an object of which that predicate is definitely true right next to one of which it is definitely false. If none of the higher-order gap principles are false, then no matter how many iterations of the definitely operator we ascend to, we will always find that between the objects that are definitely definitely^n tall and those that are definitely not definitely^n tall are some objects; but since not all of the higher-order gap principles are true, we will not always be able to *truly* describe these objects as neither definitely nor definitely not definitely^n tall. We can represent the situation in a diagram:



Given *D*-introduction, those objects that are definitely^n tall are precisely the objects that are definitely^{n+1} tall. However, the envisaged truth-value gap theorist does not always accept arguments by *reductio* or contraposition of implications. So in particular, those objects that are not definitely^n tall may be properly included among those objects that are not definitely^{n+1} tall. (For example, the things that are definitely not tall, and also the things that are borderline tall, are among the things that are not definitely tall.) But since our series is finite, this *proper* inclusion cannot continue indefinitely. But still, if no higher-order gap principle is *false*, then for any *n* there will still be a gap between the objects of which 'definitely^n tall' is definitely true and those of which it is definitely false.

This may count as infinitely higher-order vagueness. But what it really brings out is that there is no connection between infinitely higher-order vagueness and the presence of 'fuzzy boundaries', properly construed. We could have infinitely higher-order vagueness exhibited on a series of three objects: let 'definitely^n tall' be true of the first object, for any *n*; false of the last object, for any *n*; and let the middle object be such that for any *n*, we speak neither truly nor falsely when we ascribe 'definitely^n tall' to it. Then we have infinitely higher-order vagueness in the sense that, for any *n*, the segment of objects of which 'definitely^n tall' is definitely true is not contiguous with the segment of objects of which it is definitely false, and in that sense there is no sharp boundary (there being no boundary at all) between the two segments of objects. But there is no sense in which it would be correct to say that there *is* a fuzzy boundary between these two classes of objects.

2. How Classical is the Supervaluationist's Logic?

I turn now to a different set of problems, directed at supervaluationists in particular, that are raised by gap principles formulated as arguments.

$$\frac{\Phi(x) \wedge \neg\Phi(y)}{\therefore \neg R(x, y)}$$

Ordinary argumentation makes use of a notion of consequence with distinctively classical properties. So, one ordinary way to argue for a conditional is to assume the antecedent and argue for the consequent from that assumption. It is well known, however, that argument by conditional-introduction for example, or by contraposition or *reductio ad absurdum*, is not acceptable given certain treatments of vagueness involving truth-value gaps. For those who retain truth-functionality for the sentential connectives and quantifiers, no classically valid formula is deemed valid, since any such formula may have only 'indefinite' constituents (or instances, in the case of quantified formulas). Suppose, for example, that the sentence 'Al is tall' is indefinite, but that 'Al is bald' is false. Then given Kleene's strong three-valued truth tables (as adopted, for example, by Soames 1999, Tappenden 1993, and Tye 1994) the classically valid conditional 'If Al is tall then he is either bald or tall' will itself be indefinite. But still, the antecedent of this conditional is taken to *imply* its consequent in the sense that it is impossible for the antecedent to be true while the conclusion is not true. So conditional-introduction (if $\Gamma, \varphi \models \psi$

then $\Gamma \models \varphi \rightarrow \psi$) fails. Supervaluationists, who eschew truth-functionality along with bivalence, typically claim to have the upper hand here. For them, the conditional 'If Al is tall then he is either bald or tall' is true (indeed, valid) since classically true for every way ('admissible' or not) of making the vague predicates in it completely precise. The conditional is classically true on any precisification that falsifies its antecedent; and classically true also on any precisification that verifies its antecedent, since any such precisification also verifies its consequent.

Supervaluationists cannot endorse every case of argumentation by conditional-introduction, however, but it has been suggested that this restriction on our ordinary deductive practice is acceptable, because it is limited to arguments involving the definitely operator, or similar constructions. According to Fine (1975: 290), the failure of conditional-introduction 'distinguishes the presence of *D* from its absence', and according to Keefe (2000: 178) 'the cases in which [conditional-introduction and other classical principles] fail all involve the *D* operator (or similar such devices)'. While in one sense these claims are incontrovertible, there is another sense in which they are false. When relations of consequence other than the rather stringent logical consequence relation are at issue, the supervaluationist must countenance failures of conditional-introduction even in the absence of a definitely operator or any similar such devices.

The question whether supervaluationists *really* endorse classical logic has received a fair amount of attention since Williamson's (1994, ch. 5) discussion of it. The reason some commentators have claimed that the supervaluationist's semantics for a language with a definitely operator yields something less than classical logic is not that there turn out to be classically valid formulas or even arguments that are not supervaluationally valid, but rather that the supervaluationist's semantics for the language with the definitely operator yields a consequence relation that isn't closed under such classical operations as conditional-introduction, contraposition (if $\Gamma, \varphi \models \psi$, then $\Gamma, \neg\psi \models \neg\varphi$), or reductio ad absurdum (if $\Gamma, \varphi \models \psi \wedge \neg\psi$, then $\Gamma \models \neg\varphi$). For example, on versions of supervaluation semantics that endorse *D*-introduction,[12] 'Al is bald' implies 'Al is definitely bald'; but the conditional 'If Al is bald then Al is definitely bald' need not be true—it is not true when Al is a borderline case. My focus in this section is on the question whether the classical closure principles fail *only* in the presence of the definitely operator.

[12] Fine (1975); Keefe (2000).

Vann McGee and Brian McLaughlin (forthcoming) have recently contributed to the debate by arguing that commitment to classical logic requires only that discharge of premises be permissible in the case of arguments that are classically valid. No argument whose validity results from the special logic of the definitely operator—or their 'determinate truth' predicate, or any other expression not traditionally treated as a logical constant—is required to support conditional-introduction.[13]

I disagree with their general claim. We who accept classical logic accept forms of argumentation involving conditional-introduction and contraposition irrespective of whether the consequence relation at issue is logical consequence or just our everyday context-dependent conception of what follows from what. For example, if you accept that it in some sense follows from someone's being a friend of yours that you have an obligation to convey your condolences to him if he has recently suffered the death of a loved one, then you must—it seems to me—accept that in the same sense it follows from your having no obligation to convey condolences to someone who has recently suffered the death of a loved one that he is not a friend of yours.

Argument IA

> *X* is a friend of yours.
> Therefore, if *X* has recently suffered the death of a loved one, then you have an obligation to convey your condolences to *X*.

Argument IB

> *X* has recently suffered the death of a loved one.
> You have no obligation to convey condolences to *X*.
> Therefore, *X* is not a friend of yours.

I do not suggest that you must accept either argument; you might think that you incur an obligation to convey condolences to your bereaved friend only once you *know* about his loss. But that is not to object to my point, which is rather that *if* you accept that in the A case the conclusion follows from the premises then you must accept it in the B case as well. The converse also holds. The two arguments stand or fall together.

Similarly, on the classical perspective, since Sorites paradoxes show us that there is no sense in which it follows from *X*'s being tall and *X*'s being no more

[13] McGee and McLaughlin have a stake in the issue, since even though they do not endorse *D*-introduction, they are still committed to failures of conditional-introduction in the presence of the *D* operator, as pointed out by Williamson (forthcoming).

than a millimetre taller than Y that Y is tall as well,[14] it also shows us that there is no sense in which it follows from X's being tall and Y's not being tall that X is more than a millimetre taller than Y.

Argument IIA

X is tall.
X is no more than a millimetre taller than Y.
Therefore, Y is tall.

Argument IIB

X is tall.
Y is not tall.
Therefore, X is more than a millimetre taller than Y.

Supervaluationists, however (and other truth-value gap theorists), do think that there is at least some important sense in which in case IIB we have a genuine entailment. For they think that on a Sorites series for the predicate 'tall', there are borderline cases between those of which the predicate is true and those of which it is false. Argument IIB is an instance of the argument-schema formulation of a gap principle. No definitely operator is involved, but contraposition or conditional-introduction on the argument-schema formulation of the gap principle yields a Sorites paradox.

Background

Since I want in this section to spell out exactly what's meant in saying that the supervaluationist does or does not endorse classical logic, it will help for the sake of concreteness to include a brief presentation of a supervaluation semantics.[15] It will be important for what follows that the language for which the semantics is given is a language whose predicates are expressions of English. A supervaluational model ('model' for short) is a *specification space* consisting of one or more *specification points*, each of which may be thought of as a classical model, including a domain of discourse and an assignment of

[14] Richard Feldman has pointed out to me that I should qualify my claim here, since there are inductive or probabilistic notions of 'follows from'. When I say that there is some sense in which B follows from A, I mean that there is some sense in which the truth of A *guarantees* the truth of B.

[15] My presentation is based on Fine's.

extensions from that domain to the predicates in the language.[16] A formula is true in a model just in case it is true at every point in the model; false in a model just in case it is false at every point in the model.[17] Truth value at a point for formulas is defined with the standard classical truth clauses, with the result that at any given point in a model every formula is either true or false (not both).[18] A formula is neither true nor false in a model when it is true at at least one point in the model and false at at least one other.

So what happens when a definitely operator is added to the language? As before, any formula, including one containing D, is true in a (supervaluational) model just in case it is true at every point in the model. What needs to be added is a clause for truth-at-a-point for D-initial formulas. One might like to say, as Fine provisionally does, that a formula $D\varphi$ is true at a point in a model just in case φ is true at every point in the model. Adoption of this truth clause, however, precludes the possibility of there being borderline cases at even the second order (e.g. borderline definite cases or borderline borderline cases), since it makes it impossible for the value of a D-initial formula to vary from point to point within a model with the result that D-initial formulas are bivalent. To avoid this we first enrich the structure of the models by introducing an accessibility relation on the points in the model.[19] A formula $D\varphi$ is true at a point in a model just in case φ is true at every point in that model accessible to it, false otherwise.[20] (Truth-at-a-point remains bivalent.) One may add further constraints on the accessibility relation, depending on one's view of the logic of the D operator. Minimally one should require that the accessibility relation be reflexive, in order to ensure the validity of $D\varphi \rightarrow \varphi$. Nothing here will turn on any particular

[16] Perhaps one would want to allow domains to vary from point to point within a space; perhaps not. Nothing here turns on that decision, but for simplicity I assume invariant domains within a space.

[17] In the main text I leave implicit the relativization of truth in a model to a variable assignment. A formula with free variables is true (false) in a model on a variable assignment just in case it is true (false) at every point in the model on that variable assignment.

[18] The models here are structurally simpler than those in Fine's work, where models include an 'extension' relation on points and allow for there to be points at which evaluation is not bivalent. Fine's richer semantic framework allows for semantic theories other than supervaluationism, such as 'bastard intuitionism'. The simplifications here are appropriate for the supervaluation theory advocated in Fine's (1975) paper.

[19] Not to be confused with the 'extension' relation on points in Fine's models for a D-less language. See Fine (1975: 293–4) for his accessibility-relation semantics for D.

[20] $D\varphi$ is true at a point on a variable assignment, just in case φ is true at every accessible point on that same variable assignment.

properties required of the accessibility relation, though I make the minimal assumption that it is reflexive.

Treating truth as the sole designated value, a relation of logical consequence is defined as follows: an argument from a premise set of formulas Γ to a conclusion formula φ is supervaluationally valid just in case every (supervaluational) model in which all members of Γ are true is a model in which φ is true. A formula is valid when it is the conclusion of a valid argument with an empty premise set, that is, when it is true in every model whatsoever.

In what sense does the supervaluationist's semantics for the extended language yield (or not) a classical logic? At this point I prefer to rephrase the question by asking in what sense the semantics yields a classical consequence relation. In order for a consequence relation to be a classical one, it must have certain properties. It must, for example, be reflexive and transitive. It must also contain certain arguments in its extension, for example, any argument with an instance of excluded middle as conclusion. Supervaluation semantics for a standard first-order, *D*-less language yields a consequence relation that coincides in extension with the classical consequence relation. The semantics for the extended language, however, yields a consequence relation possessing some of the classical properties but lacking some others. Any classically valid argument or formula (including, e.g., $D\varphi \vee \neg D\varphi$) is supervaluationally valid.[21] Others of the classical properties of the consequence relation are guaranteed by the given model-theoretic characterization of consequence. Examples are reflexivity and transitivity, and, more generally, reiteration (if $\varphi \in \Gamma$ then $\Gamma \models \varphi$), and generalized transitivity (if $\Gamma \models \varphi$ and $\Delta \models \gamma$ for every $\gamma \in \Gamma$ then $\Delta \models \varphi$), hence also monotonicity (if $\Gamma \models \varphi$ and $\Gamma \subseteq \Delta$ then $\Delta \models \varphi$) and cut (if $\Gamma \cup \Delta \models \varphi$ and $\Gamma \models \delta$ for every $\delta \in \Delta$ then $\Gamma \models \varphi$).

But, now leading up to the main point, other classical properties are lost.[22] Examples already mentioned were failure of closure under contraposition, conditional-introduction, and *reductio ad absurdum*. To the extent that the supervaluationist wants to endorse such standard forms of argumentation under the heading 'classical logic', her aim of preserving classical logic falls short of its mark. Nevertheless, if failures of the classical closure principles

[21] The converse does not hold; since *D* is being treated as a logical constant, there are arguments (e.g. from φ to $D\varphi$) and formulas (e.g. $\neg\varphi \rightarrow \neg D\varphi$) that are supervaluationally valid yet not classically valid.

[22] Williamson (1994) and Keefe (2000) discuss this point in detail. It is also discussed by Machina (1976).

always involved the definitely operator then, as Williamson (1994: 152) puts it, 'our deductive style might not be very much cramped'. I proceed by proving, first, that failures of conditional-introduction do all involve the definitely operator when the consequence relation involved is the supervaluationist's strict relation of logical consequence. I then go on to argue that the result must be qualified and that in an important sense the supervaluationist is committed to more rampant failures of conditional-introduction.

Suppose that a formula ψ is true in every supervaluational model in which φ and all members of premise set Γ are true, i.e. that $\Gamma, \varphi \models_{SV} \psi$. Suppose also that there is a supervaluational model $\mathcal{M}$ in which every member of Γ is true, but in which $\varphi \rightarrow \psi$ is not true, i.e. that $\Gamma \not\models_{SV} \varphi \rightarrow \psi$. We can prove that some formula involved in the argument contains *D*. In the model $\mathcal{M}$ there is at least one point *p* at which $\varphi \rightarrow \psi$ is false, hence at which φ is true and ψ is false. Every member of Γ is true at *p* in $\mathcal{M}$. Now let $\mathcal{M}'$ be the model that has *p* as its only point. (There is only one such model given the assumed reflexivity of the accessibility relation.) Suppose that *D* occurs nowhere in Γ or in φ. Then φ and all members of Γ must have the same truth value at *p* in $\mathcal{M}'$ that they have at *p* in $\mathcal{M}$ since the truth-at-a-point clause for *D* is the only one that makes reference to other points in a model. So φ and all members of Γ are true at *p* in $\mathcal{M}'$, so they are true in $\mathcal{M}'$ since *p* is the only point in $\mathcal{M}'$. Since $\Gamma, \varphi \models_{SV} \psi$, ψ is also true in $\mathcal{M}'$, hence true at *p* in $\mathcal{M}'$. Since ψ has a different truth value at *p* in $\mathcal{M}'$ from the value it has at *p* in $\mathcal{M}$, it contains *D*.[23]

More Inclusive Consequence Relations

The consequence relation so far under discussion has been a relation of *logical* consequence characterized model-theoretically: in order for some premises to *logically* imply a conclusion it must be that *every* model in which the premises are all true is one in which the conclusion is as well. But among all the models are some seemingly very strange ones, since the language with which we are working is a so-called 'interpreted' one containing expressions

[23] Were there other expressions in the language, such as a truth predicate, or perhaps an '-ish' modifier, that had truth-at-a-point clauses that made reference to other points in a model, we could conclude only that the argument in question involved at least one such expression. This is what Keefe means by saying that the classical closure principles fail only in the presence of *D* 'or similar such devices' (Keefe 2000: 178).

of English. For example, there are supervaluational models that consist of exactly one point, at which 'Tall' and 'Short' have exactly the same non-empty extension: there are, for example, models in which 'Tall(Al)' and 'Short(Al)' are both true. '¬ Short(Al)' is not a logical consequence of 'Tall(Al)'. But arguments such as 'Tall(Al), therefore ¬ Short(Al)' are ones that the supervaluationist wants in some sense to validate, since they express 'penumbral connections' between the two vague predicates 'tall' and 'short'. Similarly, there are models in which an ordered pair in the extension of 'Shorter' at every point in the model has a first member in the extension of 'Tall' at every point in the model, but a second member in the extension of 'Tall' at no point in the model: there are, for example, models in which '(Tall(Al) ∧ Shorter(Al, Joe)) ⟶ Tall(Joe)' is false. The sentence is not *logically* true, but the supervaluationist wants in some sense to validate it. Preservation of penumbral connections is supposed to be one of the main advantages of the theory.

We don't yet have a problem for the supervaluationist, just the makings of one. The supervaluationist might try to forestall the problem by claiming that sentences expressing penumbral connections aren't in any sense supposed to come out *valid* on supervaluation theory; they are merely supposed to come out true as a matter of fact. Among all the models there are is one we might call the *correct* (or 'intended', or 'actual') model. The correct model is the one in which sentences have the truth value that they actually have.[24] (Correctness must probably be relativized to times and contexts of utterance.) Since I am a short philosopher, the sentence '∃x (Short (x) ∧ Philosopher (x))' is true in the correct model, though false in many others. The collection of classical models that make up the points in the correct model are those that represent the extensions our predicates would have upon admissible precisification. On no admissible precisification does a pair in the extension of 'Shorter' have a first member in the extension of 'Tall' unless its second member is as well. In the correct model, the penumbral sentence '∀x∀y((Tall(x) ∧ Shorter (x,y)) ⟶ Tall(y))' is true. This alone constitutes an advantage for supervaluations over a three-valued truth-functional theory on which, due to the presence of borderline cases, the sentence is actually neither true nor false. But still, in addition to desiring a theory on which sentences expressing penumbral connections all come out true, the super-

[24] One way that a supervaluationist might deal with higher-order vagueness would be by saying that although it is true that there is exactly one correct model, there's no model such that it is true that *it* is the correct model.

valuationist presumably also desires a theory on which *arguments* expressing penumbral connections ('Al is tall, therefore not short') come out as good in some sense. Saying that an argument expressing penumbral connection has the property that its conclusion is true in the correct model if its premises are is not saying much. Many *bad* arguments have that property since any argument with a false premise has that property.

Instead, the relevant notion of goodness may be characterized as a *relativized* consequence relation. In addition to the supervaluationist's logical consequence relation $\models_{SV}$, there corresponds to each class C of supervaluational models a relativized supervaluational consequence relation $\models_{SVC}$ defined as follows: $\Gamma \models_{SVC} \varphi$ just in case φ is true in every model in C in which all members of Γ are true. '¬ Short(Al)' is a consequence of 'Tall(Al)' relative to the class of supervaluational models that verify every penumbral sentence. My point is that when we consider consequence relations that better capture what the supervaluationist regards as a good argument than the logical consequence relation does, failures of conditional-introduction become far more rampant for the supervaluationist than is typically acknowledged. It's not that logically valid arguments are sometimes regarded as bad (except perhaps in the sense, not at issue here, of being boring or trivial), but rather that some good arguments, such as those expressing penumbral connections, are not logically valid. I do not intend to single out any one relativized consequence relation that is to be privileged as capturing, context-invariantly, a supervaluationist conception of what constitutes a good argument, or of what 'follows from' what. (We may, adapting an idea of Robert Stalnaker's, think of the class of models to which we relativize our consequence relation as growing or shrinking with the context.) What I intend to show is rather that for at least one supervaluationist conception of what follows from what that can be captured as a relativized consequence relation, conditional-introduction fails in the absence of a definitely operator.

Let me proceed by substantiating the point first informally. Supervaluationists, along with many other philosophers who think that bivalence should be rejected in the face of vagueness, think that on a Sorites series for a vague predicate, we never find an object of which the predicate is true adjacent in the series to an object of which the predicate is false. Presumably they do not think this to be the case merely as a matter of contingent fact but rather as matter of some kind of necessity, as the kind of thing you can figure out while sitting thinking in your office. From the supervaluationist's perspective, this may be expressed by saying, for example, that it's not possible for it to be true that X is tall and Y isn't unless it's true that they're not adjacent members of a Sorites series for 'tall'—it *follows from* X's being tall and Y's not being tall that they differ in height by *more* than a millimetre. (Pick a lesser difference if you think it required.) I hesitate somewhat to describe the entailment as an analytic, conceptual, or *a priori* one, since it may take some experience interacting with objects and people, reading books, or using rulers to know that *differing in height by no more than a millimetre from* is a Sorites relation for 'tall'. Nevertheless, just to have a suggestive term, I'll call the entailment an *a priori* one: according to the supervaluationist, X's being tall together with Y's not being tall *a priori* entails that X is more than a millimetre taller than Y.

But now consider the *conditional* 'If X is tall and Y isn't, then X is more than a millimetre taller than Y'. This conditional need not be true on the supervaluationist's view; though never false, it is not true when its consequent is false, X is taller than Y, and at least one of X and Y is a borderline case. One way to see the point is to note that the conditional is classically and hence supervaluationally equivalent to the conditional 'If X is tall and X is no more than a millimetre taller than Y, then Y is tall'. The thought that every such conditional is true is precisely what leads to the Sorites paradox. On the assumption that the actual heights of people form a sufficiently smooth curve, supervaluationists and epistemicists agree that at least one such conditional must be actually untrue.[25]

The relevant formal difference between supervaluation semantics and classical semantics that underlies the point is the following. Given the *classical* characterization of a model (a domain of discourse, an assignment of extensions to predicates, etc.) and of truth in a model, one may associate with each class C of classical models a relativized classical consequence relation $\models_{CLC}$ defined as in the supervaluational case: $\Gamma \models_{CLC} \varphi$ just in case φ is true in every model in C in which all members of Γ are true. The classical relation of logical consequence $\models_{CL}$ is the least inclusive relativized consequence relation, by which I mean the relation obtained by relativizing to the largest class of classical models. Not every class of classical models yields an interesting consequence relation, but since our language is an 'interpreted' one, many classes do yield an interesting consequence relation, such as the class of models that verify all analytic truths, or all sentences expressing metaphysical necessities or all those expressing *a priori* truths. In the classical case, *all* of the

[25] *Pace* Kamp (1981).

relativized consequence relations, including the uninteresting ones relativized to gerrymandered classes of models, are closed under contraposition and conditional-introduction. In the supervaluational case, however, this is not so, even if we revert to the basic semantics for the D-less language. The point can be illustrated simply by considering an uninteresting consequence relation relativized to a class C containing just one supervaluational model in which φ is neither true nor false but in which ψ is false. In the model, φ must be true at at least one point; but ψ is false at that point (since false in the model) so $\varphi \rightarrow \psi$ is false at that point and so not true in the model: $\not\models_{SVC} \varphi \rightarrow \psi$. But still, $\varphi \models_{SVC} \psi$: φ is not true in any model in C, so vacuously, ψ is true in every model in C in which φ is true. So $\models_{SVC}$ is not closed under conditional-introduction. Nor is it closed under contraposition: Although $\varphi \models_{SVC} \psi$, $\neg\psi \not\models_{SVC} \neg\varphi$, since $\neg\psi$ is true in the one model in C but $\neg\varphi$ is not.

In order to substantiate the main point of this section, I must describe a class of supervaluational models that yields an 'interesting' relativized consequence relation for which conditional-introduction fails. Building on my informal presentation of the point, let's take the relevant class C to contain those models that faithfully represent everything the supervaluationist takes herself to know *a priori*; more specifically, let's take C to contain those models that faithfully represent everything that the supervaluationist takes herself to know *a priori* about a Sorites series. When F is a vague predicate, and R is a Sorites relation for that predicate, no model that verifies both of $F(x)$ and $R(x, y)$ but falsifies $F(y)$ is in the class C. On the simplifying yet harmless assumption that R is not vague, so that $R(x, y)$ is false in any model in C in which it is not true, the relation $\models_{SVC}$ is not closed under conditional-introduction: $F(x) \wedge \neg F(y) \models_{SVC} \neg R(x, y)$; yet $\not\models_{SVC} (F(x) \wedge \neg F(y)) \rightarrow \neg R(x, y)$, since there will be models in C in which $R(x, y)$ is true, yet which contain points at which $F(x) \wedge \neg F(y)$ is true. In the familiar language of precisification: when $R(x, y)$ is true, $F(x) \wedge \neg F(y)$ can be true on a precisification, but not on every precisification.

Phenomenal Sorites

Of special interest are instances of my main example that are connected to phenomenal versions of the Sorites paradox—cases where F is a vague observational predicate, such as 'looks red' or 'tastes sweet', and where R is an associated observational sameness relation, such as the relation of *looking the same as* or *tasting the same as*. Normally, one would say that it *follows from* one thing's looking red and another thing's not looking red that the two things do not look the same; that it follows from one thing's tasting sweet and another thing's not tasting sweet that the two things do not taste the same. Conditional-introduction might get one into trouble here, however—that is, if one accepts something that few philosophers would deny, namely, that a thing that does look red and a thing that doesn't look red can be members of a sequence of things each of which looks the same as the next. For then if by conditional-introduction one concludes that if X looks red and Y doesn't then X and Y do not look the same, one can further conclude, by classical principles that the supervaluationist accepts, that if X looks red and X does look the same as Y then Y looks red as well. Repeated application of this principle along a Sorites series of the described kind results in contradiction. The last thing both does and does not look red.

The argument 'This looks red and that doesn't; therefore this does not look the same as that' has some features worth noting here. First, an epistemicist who accepts what I'll call a 'chaining principle'—that a thing that looks red can be connected by a looks-the-same-as chain with a thing that does not look red—cannot accept the argument.[26] Such a person, in keeping with his general treatment of the Sorites paradox, will say that the following universal generalization is *false*: $\forall x \forall y$ ((x looks red $\wedge$ x looks the same as y) $\rightarrow$ y looks red); and will hence say that it has a particular false instance, that for some particular values of x and y, 'x looks red' and '$\neg y$ looks red' and 'x looks the same as y' are all true. On this view the argument 'This looks red and that doesn't, therefore this does not look the same as that' is no good. (I, being an epistemicist who does accept the argument, reject the chaining principle.[27]) Second, and relatedly, the *supervaluationist* who accepts the chaining principle is committed to accepting the argument as a special case of her view that verifiers and falsifiers do not sit next to each other on a Sorites series since they are separated by intervening borderline cases. Third, unlike non-phenomenal instances of my main example, the phenomenal instances of the argument seem to be cases of arguments expressing penumbral connections. The argument 'This candy tastes sour and that candy

[26] I thank Alexis Burgess for this forceful formulation of the problem.

[27] See my 'Phenomenal Continua and the Sorites' (Graff 2001) for a defense. Any epistemicist who thinks there are non-epistemic senses of 'looks' should wholeheartedly welcome the arguments of that paper.

doesn't; therefore the two candies do not taste the same' has a conclusion that follows from its premise in just the same sense that it follows from one woman's being tall and another's not being tall that they are not the same height. But unlike the other penumbral argument we considered—Al is tall, therefore not short—this argument will not on the supervaluationist's view support conditional-introduction: the associated conditional 'If this tastes sour and that doesn't then they do not taste the same' cannot be admitted as a *sentence* expressing penumbral connection, unless of course the chaining principle is, as I think, false.

Implicit Premises

An important question emerges at this point. Can a relativized consequence relation always be alternatively characterized in supervaluation theory as logical consequence given some (perhaps infinite) set of implicit premises? In other words, when $\models_{SVC}$ is a relativized supervaluational consequence relation, must there be some set of formulas Σ, to be thought of as implicit premises, such that $\Gamma \models_{SVC} \varphi$ if and only if $\Sigma \cup \Gamma \models_{SV} \varphi$? We know that *if* C is characterized as the class of models in which every member of a set Σ of formulas is true, then $\Gamma \models_{SVC} \varphi$ if and only if $\Sigma \cup \Gamma \models_{SV} \varphi$. The question is whether for every relativized consequence relation $\models_{SVC}$ there is some set of formulas Σ meeting the stated criterion. For example, '$\forall x(\text{Tall}(x) \rightarrow \neg\text{Short}(x))$' is a penumbral sentence, so '$\neg$ Short(Al)' can be characterized as a consequence of 'Tall(Al)' relative to the class of models in which all penumbral sentences are true, or it can be characterized as a logical consequence, given penumbral sentences (though not all of them are needed) as implicit premises. But since we have at this point seen at least one example of an argument expressing penumbral connection for which conditional-introduction fails, the question arises whether the supervaluationist's penumbral consequence relation can be alternatively characterized as logical consequence plus implicit premises.

It cannot be assumed on the supervaluationist picture that every relativized consequence relation can be alternatively characterized as logical consequence plus implicit premises. In fact, if we restrict ourselves to the basic supervaluational semantics for a standard first-order language without a definitely operator, we find that we have already shown that there is a relativized supervaluational consequence relation that cannot be so characterized. We have shown that even in the absence of D, there is a class C of supervaluational models that yields a relativized consequence relation not close under conditional-introduction, so that for some Γ, φ, and ψ: $\Gamma, \varphi \models_{SVC} \psi$ but $\Gamma \not\models_{SVC} \varphi \rightarrow \psi$. But we have also seen that in the absence of D, if $\Gamma \cup \Sigma, \varphi \models_{SV} \psi$ then $\Gamma \cup \Sigma \models_{SV} \varphi \rightarrow \psi$, for all Σ.[28]

I say all this as a preface to offering a reply to the main point of this section on behalf of the supervaluationist. I argued that the relativized supervaluational consequence relation plausibly called supervaluational *a priori* consequence is one for which conditional-introduction fails even in the absence of the D operator. I argued also that if (contrary to my own view) some chaining principles are true, then even the supervaluationist's *penumbral* consequence relation is one for which conditional-introduction fails even in the absence of the D operator. In substantiating the point I characterized a class of supervaluational models in a way that made essential use of the supervaluationist's metalinguistic *truth* predicate. Noting this, the supervaluationist might like to respond to my point by arguing that the relativized *a priori* or penumbral consequence relations *can* be alternatively characterized as a relation of logical consequence plus implicit premises that do contain the D operator. Given the supervaluationist's extended language and semantics, the class of models appealed to in substantiating my main point can be characterized as the class of models in which the universal generalization of every formula $(DF(x) \wedge D\neg F(y)) \rightarrow \neg R(x, y)$ is true whenever F is a vague predicate and R stands for a Sorites relation for that predicate. So, the reply goes, the case I offered of a relativized consequence relation for which conditional-introduction fails even in the absence of the D operator was really just a case of the failure of logical consequence to be closed under conditional-introduction in the presence of the D operator, with the premises of the argument containing D serving as 'implicit' premises. The reply is to the point, but serves as something of a double-edged sword for the supervaluationist. If every worrisome failure of conditional-introduction and contraposition is to be accounted for as a case implicitly involving the D operator, then argumentation involving the D operator is much more common than it would at first appear.

[28] In fact, even with the D operator we can show, given the compactness theorem for the classical consequence relation, that not every relativized supervaluational consequence relation can be characterized as logical consequence plus implicit premises.

Implicit Premises and Higher-Order Vagueness

But there is a sharper edge to the sword: the implicit premises now being appealed to are universal closures of gap principles, which as we saw in the first section cannot all be true. Before elaborating this point, and spelling out the dilemma it poses for the supervaluationist, we should first take stock of the dialectic so far. The supervaluationist's relation of logical consequence, when expanded to capture the validity of arguments resulting from the logic of the definitely operator, does not satisfy certain classical closure principles such as closure under conditional-introduction. Failures of conditional-introduction are nevertheless, it is claimed, suitably restricted to arguments involving the definitely operator.[29] I showed that supervaluationists' own view of what constitutes a lack of sharp boundary on a Sorites series commits them to failures of conditional-introduction, even in the absence of a definitely operator, once we consider consequence relations that are less restrictive than logical consequence—such as the relation I (somewhat tendentiously) called *a priori* consequence; and also, if chaining principles are true, the relation of penumbral consequence so dear to the supervaluationist's heart. Examples, respectively, were the arguments 'Al is tall and Joe isn't, therefore they differ in height by more than a millimetre' and 'This looks red and that doesn't, therefore they do not look the same'. Since these are the sorts of consequence relations often involved in ordinary argumentation, the supervaluationist who, as part of her advocacy of classical forms of argumentation, wants failures of the classical closure principles to be suitably restricted now has something to worry about. The more inclusive consequence relations were characterized model-theoretically as relativized consequence relations for it which it can be shown that conditional-introduction fails in absence of the definitely operator. On behalf of the supervaluationist, I proposed an alternative characterization of these inclusive consequence relations as logical consequence plus sets of implicit premises, and explained why, so characterized, the definitely operator proved to be involved after all—in the set of implicit premises. It emerged that the crucial implicit

[29] Something stronger may be said: a valid argument will support conditional-introduction not only when it does not contain D, but also when it is classically valid. The classically valid argument from $D\varphi$ to $D\varphi \vee D\neg\varphi$ does support conditional-introduction. Something even stronger may be said, though I won't provide a general formulation here, since the supervaluationally valid argument from $D\varphi$ to φ also supports conditional-introduction.

premises were universal closures of gap principles (conditional formulation). If the supervaluationist is to appeal to such sentences as implicit premises, then she had better think that they are true. But as we saw in the first section, she cannot think they are *all* true, on pain of contradiction.

I'll close by posing the following dilemma for the supervaluationist who accepts the validity of *D*-introduction. If she thinks that there is infinitely higher-order vagueness, in particular if she thinks that for any n, there is a gap on a Sorites series (even one of finite length) between the things of which $D^nF(x)$ is definitely true and those of which it is definitely false, then she is committed to the validity (in some inclusive sense) of the argument from $D^nF(x)$ and $\neg D^nF(y)$ to $\neg R(x, y)$, for any n. This is an argument for which conditional-introduction fails, and its validity cannot be alternatively characterized as logical consequence plus implicit premises the supervaluationist accepts as true, since we saw at the outset that for some m, the corresponding gap principle is not true. In the case where $n = 0$, we have an argument for which contraposition and conditional-introduction fail even in the absence of a definitely operator. And since n^{th}-order vagueness turns out not to be so intimately connected with n^{th}-order gap principles, it is unclear why the supervaluationist should think that first-order vagueness is best expressed by a first-order gap principle. For, as we saw, n^{th}-order vagueness can be secured by the non-falsity of an n^{th}-order gap principle, which on the view under discussion cannot be equated with its truth. If the supervaluationist does not think that there is infinitely higher-order vagueness, in particular, if for some n, she accepts a sharp boundary (no gap) between the things of which $D^nF(x)$ is true and those of which it is false, then it is unclear what her motivation is for rejecting such a boundary at the first level.

REFERENCES

EDGINGTON, D. (1993), 'Wright and Sainsbury on Higher-Order Vagueness', *Analysis*, 53/4: 193–200.

FINE, K. (1975), 'Vagueness, Truth and Logic', *Synthese*, 30: 265–300.

GÓMEZ-TORRENTE, M. (1997), 'Two Problems for an Epistemicist View of Vagueness', in E. Villanueva (ed), *Philosophical Issues*, viii: *Truth* (Atascadero, Calif.: Ridgeview).

——(2002), 'Vagueness and Margin for Error Principles', *Philosophy and Phenomenological Research*, 64/1: 107–25.

GRAFF, D. (2001), 'Phenomenal Continua and the Sorites', *Mind*, 110/440: 905–35.

——(2002), 'An Anti-epistemicist Consequence of Margin for Error Semantics for Knowledge', *Philosophy and Phenomenological Research*, 64/1: 127–42.

——and T. Williamson (eds.) (2002), *Vagueness* (Burlington, Vt.: International Research Library of Philosophy, Ashgate).

HECK, R. G., Jr. (1993), 'A Note on the Logic of (Higher-Order) Vagueness', *Analysis*, 53/4: 201–8.

KAMP, H. (1981), 'The Paradox of the Heap', in U. Mönnich (ed), *Aspects of Philosophical Logic* (Dordrecht: Reidel).

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

MCGEE, V., and B. MCLAUGHLIN (1998), 'Review of Timothy Williamson's *Vagueness*', *Linguistics and Philosophy*, 21: 221–35.

—— ——(forthcoming), 'Logical Commitment and Semantic Indeterminacy: A Reply to Williamson', *Linguistics and Philosophy*.

MACHINA, K. F. (1976), 'Truth, Belief, and Vagueness', *Journal of Philosophical Logic*, 5: 47–78.

SAINSBURY, R. M. (1991*a*), *Concepts without Boundaries*, King's College London, Department of Philosophy, Inaugural lecture delivered 6 Nov. 1990.

——(1991*b*), 'Is there Higher-Order Vagueness?', *Philosophical Quarterly*, 41/163: 167–82.

SOAMES, S. (1999), *Understanding Truth* (New York: Oxford University Press).

TAPPENDEN, J. (1993), 'The Liar and Sorites Paradoxes: Toward a Unified Treatment', *Journal of Philosophy*, 90/11: 551–77.

TYE, M. (1994), 'Sorites Paradoxes and the Semantics of Vagueness', *Philosophical Perspectives*, 8: 189–206.

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

——(1997), 'Imagination, Stipulation and Vagueness', in E. Villanueva (ed), *Philosophical Issues*, viii: *Truth* (Atascadero, Calif.: Ridgeview).

——(1999), 'On the Structure of Higher-Order Vagueness', *Mind*, 108/429: 127–43.

——(2002), 'Epistemicist Models: Comments on Gómez-Torrente and Graff', *Philosophy and Phenomenological Research*, 64/1: 143–50.

——(forthcoming), 'Reply to McGee and McLaughlin', *Linguistics and Philosophy*.

WRIGHT, C. (1975), 'On the Coherence of Vague Predicates', *Synthese*, 30: 325–65.

——(1987), 'Further Reflections on the Sorites Paradox', *Philosophical Topics*, 15: 227–90.

——(1992), 'Is Higher-Order Vagueness Coherent?', *Analysis*, 52/3: 129–39.