Williams on Supervaluationism and Logical Revisionism

Nicholas K. Jones

Non-citable draft: 26–02–2010.

Final version appeared in: *The Journal of Philosophy* (2011) 108: 11: 633-641

Central to discussion of supervaluationist accounts of vagueness is the extent to which they require revisions of classical logic and if so, whether those revisions are objectionable. In an important recent *Journal of Philosophy* article, J.R.G. Williams presents a powerful challenge to the orthodox view that supervaluationism is objectionably revisionary. Williams argues both that supervaluationism is non-revisionary and that even if it were, those revisions would be unobjectionable (Williams, 2008). This note shows that his arguments for both claims fail.

1 The case for revisionism

Williams begins with a model-theoretic characterisation of consequence in a supervaluationist setting. A *supervaluationist model* M is a quadruple $\langle D_M, \Delta_M, R_M, I_M \rangle$. D_M is the domain, Δ_M a set of delineations, R_M an accessibility relation on Δ_M, and I_M an interpretation function from expressions and delineations onto classical extensions. Truth relative to models and delineations—truth at M, d—is defined in the standard way. This much is exactly analogous to a possible-worlds semantics (without an actual world) for a modal language.

Supervaluationist semantics departs from standard modal semantics in its characterisations of truth and consequence. Say that a sentence Φ is *supertrue* in M iff Φ is true at M, d, for every delineation $d \in \Delta_M$. Truth in a model M is then identified with supertruth in M. Since some sentences are neither supertrue nor superfalse in some models, bivalence fails.

Both local and global consequence relations are definable within this framework:

$\Gamma \vDash_{\text{local}} \Phi$ iff, for every model M and delineation $d \in \Delta_M$, if every member of Γ is true at M, d, then Φ is true at M, d.

$\Gamma \vDash_{\text{global}} \Phi$ iff, for every model M, if every member of Γ is supertrue in M, then Φ is supertrue in M.

Given the identification of truth with supertruth, there are well-known reasons to identify consequence proper with global consequence, rather than its local counterpart (see, for example Williamson, 1994, p.148). To these, Williams adds a compelling new argument (§3). So let us set aside local consequence: so far as supervaluationism is concerned, global consequence is consequence.

The purported counterexamples to classical logic arise with the introduction of a 'Definitely' operator D, akin to a necessity operator in possible-worlds semantics:

$D\Phi$ is true at M, d iff Φ is true at M, d', for every delineation $d' \in \Delta_M$ R_M-accessible from d.

The following results all hold, providing apparent counterexamples to the respective classical rules of proof:

Contraposition:

- $p \vDash_{\text{global}} Dp$
- $\neg Dp \nvDash_{\text{global}} \neg p$

Conditional proof:

- $p \vDash_{\text{global}} Dp$
- $\nvDash_{\text{global}} p \supset Dp$

Argument by cases:

- $\vDash_{\text{global}} p \vee \neg p$
- $p \vDash_{\text{global}} Dp$
- $\neg p \vDash_{\text{global}} D\neg p$
- $\nvDash_{\text{global}} Dp \vee D\neg p$

Reductio:

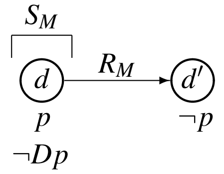
- $p \wedge \neg Dp \vDash_{\text{global}} \bot$
- $\nvDash_{\text{global}} \neg(p \wedge \neg Dp)$

None of these results holds for local consequence. So logical revisionism could be avoided by identifying that with consequence proper. But this is not Williams's approach. His strategy is to argue that the present supervaluationist framework is inadequate; the results all fail in a more satisfactory setting.

2 Williams's case against revisionism

Williams begins by observing that plausible semantic analyses of linguistic phenomena other than vagueness, specifically comparatives, have been proposed that require delineations other than those relevant to the determination of (super)truth-value (§2). A fully general semantic theory for a vague language may well have to incorporate such delineations. To accommodate this, let an *extended model* M be just like a supervaluationist model, but with a fifth element: $\langle D_M, \Delta_M, R_M, I_M, S_M \rangle$, where $S_M \subseteq \Delta_M$. The elements of S_M are the *sharpenings* of M. Supertruth in M is then redefined as truth at M,d, for all sharpenings $d \in S_M$. The sharpenings are the privileged subset of delineations that contribute to determining truth-value. Extended models are just like standard possible-worlds models, but with a set of actual worlds. If the quantifiers over models in our characterisation of global consequence range over extended models, then all the results in the previous section fail.

To see this, note first that central to the purported counterexamples are the results $p \vDash_{\text{global}} Dp$ and $p \wedge \neg Dp \vDash_{\text{global}} \bot$. Neither holds in Williams's extended setting. For let M be an extended model with just two delineations d, d', whose only sharpening is d, where d R_M-accesses d', and such that p is true at M,d but not true at M,d':



Since d is the only sharpening and p is true at M,d: p is supertrue in M. Since d' is R_M-accessible from d, and p is not true at M,d': Dp is not true at M,d. So $\neg Dp$ is true at M,d, and hence supertrue in M. So M is a countermodel to both $p \vDash_{\text{global}} Dp$ and $p \wedge \neg Dp \vDash_{\text{global}} \bot$.

The game is not yet up. Let the *admissible* extended models be those over which the quantifiers in our characterisation of global consequence range. Let the *restricted-access* (RA) models be those where the sharpenings access only other sharpenings. It is crucial to Williams's case that the admissible models not be restricted to RA-models. For suppose that p is supertrue in an RA-model M. Then no sharpening accesses any delineation where p is not true. So Dp is supertrue in M. Since M was arbitrary: $p \vDash_{\text{global}} Dp$, and hence $p \wedge \neg Dp \vDash_{\text{global}} \bot$, hold unless the admissible models include some non-RA-models. A two-step strategy for reinstating the counterexamples now emerges. First step: argue that only RA-models respect the intended sense of D. Second step: argue that admissible models must respect the intended sense of D.

Williams pre-empts this strategy. Against the first step, he constructs a toy non-RA-model which, he claims, does respect the intended sense of D (§5). He continues:

> "Perhaps some inventive defender of the orthodox position could make a case that the toy model constructed above, and all others like it, are unfaithful to the intended sense of 'Definitely' (I have no idea what such a case would look like, but cannot rule it out)." (Williams, 2008, p.205)

Section 3 provides just such a case. Against the second step, he argues that restricting the admissible models to RA-models involves treating D as a logical constant. Since the proper characterisation of the logical constants is highly contentious, Williams concludes that the case for logical revisionism is weak at best (§6). Section 4 argues that the case for logical revisionism can sidestep this issue: the proper characterisation of the logical constants is irrelevant. Section 5 closes by showing that even if Williams is right on both counts, a slightly different argument for logical revisionism is available that avoids his complaints. Furthermore, this last argument undermines Williams's case for regarding the purported revisions as unobjectionable.

3 Reinstating revisionism: first step

This section argues that supervaluationists should regard only RA-models as faithful to the intended sense of D. The argument rests on two explanatory demands on a satisfactory account of vagueness. The first is an account of borderline status and definiteness: what is it for a sentence to be definitely true, or an object to be definitely F?[1] The second is an account of borderline ignorance: if it is borderline whether p, then it seems somehow misguided to try and discover whether p. Why does borderline status interact with knowledge in this way? And if it does not, why does it appear to?

Supervaluationism offers an account of borderline status, and hence definiteness, in terms of truth-value gaps: it is borderline whether p iff p is neither true nor false (in virtue of being true at some but not all sharpenings). Thus definiteness is analysed in terms of more familiar semantic concepts. Note also that without the analysis of borderline cases in terms of truth-value gaps, the identification of truth with super-truth looks ill-motivated. For that identification is motivated by the desire to avoid bivalence; unless borderline cases fall down a truth-value gap, it is entirely mysterious why a non-bivalent semantics would be desirable. This analysis of borderline status naturally extends to an explanation of borderline ignorance: since knowledge implies truth, if p is borderline (neither true nor false), it cannot be known whether p.[2] Thus the importance that D provide "an object-language reflection

[1] Some deny that any analysis of borderline status is possible (Barnett, 2009). I take it that the lack of explanatory work for the concepts of definiteness and of borderline status outside of vagueness counts strongly against these views.

[2] Field (2008, pp.154–5) criticises this account of borderline ignorance. But without an alternative, the supervaluationist should endorse it. I see only one alternative explanation for the appearance of borderline ignorance: although borderline status does not prevent knowledge, it does prevent clear knowledge; and one seems ignorant when one does not clearly know. This fits Williams's extended semantics that permits true borderline statements. But this just strengthens the demand for an explanation of definiteness. Supposing I know that p, why does it matter whether I also definitely know that p? Why should an inability to definitely know make it (or make it appear) futile to try and discover whether p? Some account is surely owed, and it is hard to see where it might come from, if not from the analysis of definiteness. I see two options: (1) It is just a primitive fact about vagueness that borderline status makes it appear futile to try to know. (2) Definite knowledge is the goal of assertion, and borderline status creates the appearance of ignorance by making it in principle illegitimate to assert that p when p is borderline. Neither approach is satisfactory without an account of definiteness. Why is definite knowledge, rather than mere knowledge, the goal of assertion? Simply postulating primitive and inexplicable relationships between definiteness and other concepts is an unsatisfying approach to

of the supervaluationist's notion of truth—"supertruth" " (Williams, 2008, p.192).

Neither explanation is available if non-RA-models respect the intended sense of D. For in some non-RA-models, p is both true and borderline (as the model diagrammed in the previous section shows). Hence borderline status cannot be analysed in terms of truth-value gaps, and borderline ignorance cannot be explained in terms of the untruth of borderline statements, if such models are faithful to the intended sense of D. If the intended sense of 'definitely' permits these explanations, as it must if the identification of truth with supertruth is to be well-motivated, then only RA-models respect that sense.[3]

One might reply that $p \vDash_{\text{global}} Dp$ should fail anyway in the presence of higher-order vagueness. For an instance is $Dp \vDash_{\text{global}} DDp$, which rules out borderline cases to the clear cases. The objection is mistaken. If $p \vDash_{\text{global}} Dp$, then $Dp, DDP, DDDp, \ldots$ are all true in models where p is true. But this is consistent with the falsity of $Dp, DDp, \ldots$ and untruth of the S4 principle $Dp \supset DDp$ in models where p is borderline (and hence untrue). By identifying truth with supertruth with clear truth, supervaluationism collapses the clear, clearly clear, etc., cases into the cases. But it does not reduce higher-order borderline status to inconsistency, as the objection assumes.

I conclude that the supervaluationist should regard only RA-models as faithful to the intended sense of D if they are to offer their customary semantic analysis of definiteness and explanation of borderline ignorance, as well as the point of identifying truth with supertruth. Since no alternative explanation is forthcoming, the supervaluationist should regard only RA-models as faithful to the intended sense of D.

4 Reinstating revisionism: second step

Should the admissible models be restricted to those faithful to the intended sense of D? Expressions whose interpretation is held constant across admissible models when characterising Logical Consequence are called Logical Constants. So our question becomes: is D a Logical Constant? Williams argues that not only is the proper characterisation of the

discharging the supervaluationist's explanatory obligations.

[3] It does not strictly follow that only RA-models respect the intended sense of D. The argument provides only the following necessary condition for doing so: no sharpening accesses any delineation where any supertruth is not true. The restriction of admissible models to those that meet this constraint will reinstate the counterexamples to the classical rules. So we can safely ignore this complication.

Logical Constants too controversial to provide a firm basis on which to rest the case for revisionism, but even the proper application of competing accounts to particular cases is disputed. This is surely correct. So let us set aside the question of whether D is a Logical Constant. Consider instead the following questions: (1) Why are the interpretations of some expressions held fixed across admissible models? (2) What is an investigation of global consequence supposed to provide? We address each in turn.

Why hold fixed the interpretation of $\wedge$ when investigating the logical behaviour of conjunction? Why not permit models that interpret $\wedge$ as disjunction? The answer is that these deviant interpretations are irrelevant to our interests. We want to know what follows from the conjunction of p and q, not from the truth of the string of symbols '$p \wedge q$'. Regardless of whether $\wedge$ is a Logical Constant, or Logical Consequence requires constancy in its interpretation across admissible models, if we want to know how conjunction contributes to the correctness of inferences, then we should investigate what consequence relation results when $\wedge$ is interpreted as conjunction in all admissible models. The point is that deductive inference involves the manipulation of contents, not strings of symbols; it involves discovering what must be the case, given other beliefs and suppositions as to what is the case. Some consequence relation provides the standard of correctness for this activity, regardless of whether that relation is Logical Consequence proper or the inferences in question are strictly Logical. So if we want to know how to reason with conjunctive contents, we should, on pain of simply changing the topic, investigate consequence relations where $\wedge$ is interpreted as conjunction in all admissible models.

Let us turn to our second question: what should our investigation of global consequence provide? In part, we want an account of correct reasoning in a vague language. But we also want to know how to reason about vagueness: what follows from the supposition that p is borderline? The sole purpose of the D operator is to allow us to make such suppositions, to allow expression of definitised contents in a language with supervaluationist semantics. Thus our investigation of global consequence should provide an account of the contribution of D to the correctness of inferences. Classical predicate calculus and model-theory provide a powerful tool for studying reasoning with predicative, conjunctive, disjunctive, negated and quantified contents in a precise language. Supervaluationist model-theory is supposed to provide a similar tool for studying reasoning with those same kinds of content in a vague language. With the addition of D, we acquire the further capacity to study reasoning with contents concerning definiteness also.

Combining these answers to our two questions yields the following: if we want to know how we should reason about definiteness, then we should investigate consequence relations where the admissible models are faithful to the intended sense of D. It is simply irrelevant to this investigation whether D is a Logical Constant, or whether any such consequence relation is Logical Consequence. We want to know how to reason about definiteness in a vague language. Given the argument of the previous section, it follows that only RA-models are admissible. But then the initial results all hold. Each provides a counterexample to the claim that all classically correct inferential patterns are legitimate in a language with supervaluationist semantics and the resources for expressing definiteness. Whether or not this is a revision to Logic, the supervaluationist should not reason classically about definiteness.

5 A better case for revisionism

Suppose that my arguments in the preceding sections all fail. Then because not all admissible models are RA-models: $p \nvDash_{\text{global}} Dp$ and $p \wedge \neg Dp \nvDash_{\text{global}} \bot$. We can show that supervaluationism is revisionary nonetheless. The reason is that nothing prevents the introduction of a truth-operator via the following clause:

$\mathsf{T}p$ is true at M, d iff p is true (i.e. supertrue) in M.

The initial results all hold under a uniform substitution of D for T: T induces new counterexamples to the classical rules. Since the semantic clause for T is given in terms of truth in a model directly, there is no scope to avoid this by tinkering with the underlying space of sharpenings or an accessibility relation on them. Setting disputes about D to one side, the equivalence of $\mathsf{T}p$ to the truth of p at all and only the sharpenings is mandated by the identification of truth with supertruth and adoption of a global consequence relation. The new counterexamples to the classical rules thus follow directly from the key supervaluationist claim to retain LEM without Bivalence. Furthermore, it is in terms of truth, rather than definiteness, that much discussion of the revisionary implications of supervaluationist semantics has been conducted (McGee & McLaughlin, 1998, 2004; Williamson, 2004).[4]

[4] In an appendix, Williams offers a proof that global and classical consequence coincide. It is invalid for languages containing T. To show this, first we need an account of classical consequence. A standard possible-

One might object that Liar-like paradoxes provide independent reason to doubt that the classical rules hold in full generality anyway, for languages that contain their own truth-predicate. But since $\top$ is not a predicate, but a sentential operator, we cannot use it to construct a "Liar sentence" or similar. The objection therefore fails.

I close by noting that $\top$ undermines Williams's case against regarding deviations from the classical rules as revisionary to classical inferential practice, as opposed to classical logical theory. Let $\vDash_{\text{global}}$ be the global consequence relation obtained by allowing admissible non-RA-models; let $\vDash^{+}_{\text{global}}$ be the global consequence relation obtained by restricting admissible models to RA-models. Williams argues that $\vDash^{+}_{\text{global}}$ does not involve deviation from classical inferential practice by appeal to:

CP* If $\Gamma, A \vDash_{\text{global}} C$, then $\Gamma \vDash^{+}_{\text{global}} A \supset C$.

His idea is that, in order to show that supervaluationism mandates revisions to inferential practice, "we would have to show that inferential practice mandates moving from $\vDash^{+}_{\text{global}}$-valid but $\vDash_{\text{global}}$-invalid arguments to conditional conclusions. No such case has been made." (Williams, 2008, p.210) But in a language containing $\top$, even CP* fails:

$$p \vDash_{\text{global}} \top p, \text{ but: } \nvDash^{+}_{\text{global}} p \supset \top p.$$

Even if inferential practice involves no more than drawing conditional conclusions from $\vDash_{\text{global}}$-valid arguments, supervaluationism is revisionary of that practice if, as argued above,

---

worlds model for a modal language is just one of our extended models with a single sharpening @. Call any such model a *classical model*. Then we have: $\Gamma \vDash_{\text{classical}} \Phi$ iff, for every classical model M, if every member of Γ is true at $M, @$, then Φ is true at $M, @$. Williams now has to show that if $\Gamma \nvDash_{\text{global}} \Phi$, then $\Gamma \nvDash_{\text{classical}} \Phi$. He proceeds as follows:

Suppose $\Gamma \nvDash_{\text{global}} \Phi$. Then for some extended model M: every member of Γ is supertrue in M and Φ is not supertrue in M. So for some sharpening $s \in S_M$: every member of Γ is true at M, s and Φ is not true at M, s. Let M^* be the extended model that differs from M only by substituting $\{s\}$ for S_M. M^* is a classical model where $@ = s$. By construction: every member of Γ is true at $M^*, @$ but Φ is not true at $M^*, @$. So M^* is a countermodel to: $\Gamma \vDash_{\text{classical}} \Phi$. Discharging our initial supposition: if $\Gamma \nvDash_{\text{global}} \Phi$, then $\Gamma \nvDash_{\text{classical}} \Phi$.

The proof fails in a language containing $\top$. The problem is that although Γ are all true at M, s and Φ is not true at M, s, there is no guarantee that both (i) Γ are all true at M^*, s, and (ii) Φ is not true at M^*, s. To see this, let $\Gamma = \{\neg \top p\}$ and let $\Phi = \neg p$. Although there are extended models where $\neg \top p$ is true and $\neg p$ is not true, there is no classical model where this is so. For if $\neg p$ is not true at $M^c, @$ for some classical model M^c, then p is true at $M^c, @$. So $\top p$ is true at $M^c, @$. Hence $\neg \top p$ is not true at $M^c, @$. Williams proof fails because: $\neg \top p \nvDash_{\text{global}} \neg p$ but $\neg \top p \vDash_{\text{classical}} \neg p$.

$\models^{+}_{\text{global}}$ is the standard for deductive correctness when reasoning using D. Contra Williams, supervaluationism does lead to logical revisionism.

References

Barnett, D. (2009), "Is vagueness *sui generis*?" *Australasian Journal of Philosophy*, 87: 5–34.

Field, H. (2008), *Saving truth from paradox*, Oxford: OUP.

McGee, V. & McLaughlin, B. (1998), "Review of Williamson (1994)," *Linguistics and Philosophy*, 21: 221–35.

— (2004), "Logical commitment and semantic indeterminacy: a reply to Williamson," *Linguistics and Philosophy*, 27: 122–36.

Williams, J.R.G. (2008), "Supervaluationism and logical revisionism," *Journal of Philosophy*, 55(4): 192–212.

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

— (2004), "Reply to McGee and McLaughlin," *Linguistics and Philosophy*, 27: 113–22.