Journal of Philosophical Logic (2022) 51:1549–1582
https://doi.org/10.1007/s10992-021-09618-4

Meta-inferences and Supervaluationism

Luca Incurvati[1] · **Julian J. Schlöder**[1]

Received: 27 May 2020 / Accepted: 1 July 2021 / Published online: 3 August 2021

Abstract

Many classically valid meta-inferences fail in a standard supervaluationist framework. This allegedly prevents supervaluationism from offering an account of good deductive reasoning. We provide a proof system for supervaluationist logic which includes supervaluationistically acceptable versions of the classical meta-inferences. The proof system emerges naturally by thinking of truth as licensing assertion, falsity as licensing negative assertion and lack of truth-value as licensing rejection and weak assertion. Moreover, the proof system respects well-known criteria for the admissibility of inference rules. Thus, supervaluationists can provide an account of good deductive reasoning. Our proof system moreover brings to light how one can revise the standard supervaluationist framework to make room for higher-order vagueness. We prove that the resulting logic is sound and complete with respect to the consequence relation that preserves truth in a model of the non-normal modal logic **NT**. Finally, we extend our approach to a first-order setting and show that supervaluationism can treat vagueness in the same way at every order. The failure of conditional proof and other meta-inferences is a crucial ingredient in this treatment and hence should be embraced, not lamented.

Keywords Meta-inferences · Supervaluationism · Global consequence · Multilateral logic · Higher-order vagueness

1 Introduction

Necessary truth is often identified with truth in all possible worlds. The supervaluationist approach to vagueness identifies truth *simpliciter* with truth on all

✉ Luca Incurvati
l.incurvati@uva.nl

Julian J. Schlöder
julian.schloeder@gmail.com

[1] Department of Philosophy & ILLC, University of Amsterdam, Amsterdam, Netherlands

precisifications. David Lewis [17] famously said that possible worlds are ways the world could have been. Precisifications are ways language could be made precise.

In standard modal logic, consequence is defined *locally*, as the preservation of truth at a world. In a supervaluationist framework, it is natural to define consequence *globally*, as the preservation of truth on all precisifications. But if the supervaluationist object language contains a *definitely* operator Δ, certain classical meta-inferences fail for global consequence, notably conditional proof, *reductio ad absurdum*, proof by cases, contraposition and, in a first-order setting, existential instantiation. This calls into question whether supervaluationism can provide an account of good deductive reasoning [10, 11, 28]. If certain classically valid argument patterns fail, supervaluationists must indicate which argument patterns may be employed in legitimate inference. This is the *proof-theoretic problem* for supervaluationism.

To solve the proof-theoretic problem, we provide a proof system for supervaluationist logic which includes supervaluationistically acceptable versions of the problematic meta-inferences. The proof system is a multilateral calculus in the style we introduced in [13] and [14]: inferences may involve *asserted* sentences as well as *rejected* and *weakly asserted* ones. The proof system emerges naturally by thinking of truth *simpliciter* as licensing assertion, falsity *simpliciter* as licensing negative assertion, and lack of truth-value as licensing rejection and weak assertion. Moreover, the proof system respects well-known criteria for the admissibility of inference rules. The supervaluationist can specify which argument patterns govern good deductive reasoning when vague terms are involved.

According to supervaluationism, the vagueness of *bald* lies in the fact that it admits borderline cases—it may be that it is neither definitely the case that Harry is bald nor is it definitely the case that he is not bald. Now, the logic of assertion of our proof system is sound and complete with respect to the global consequence relation of the modal logic **S5**. It follows that it is either definitely the case that Harry is definitely bald or it is definitely the case that he is not definitely bald. However, higher-order predicates such as *definitely bald* would seem to admit of borderline cases, just as their first-order counterparts. This is the *basic problem of higher-order vagueness* for supervaluationism. To solve it, it is natural for supervaluationists to abandon the modal logic Axiom **4** and retreat to a modal logic in which it no longer follows that it is either definitely the case that Harry is definitely bald or it is definitely the case that he is not definitely bald.

However, the broader problem of higher-order vagueness for supervaluationism cuts deeper than this. Delia Graff Fara [10] argued that to account for higher-order vagueness, the supervaluationist must accept *gap principles* stating that a definitely bald person cannot become not definitely bald by gaining one hair, and similarly for any finite iteration of *definitely*. The gap principles give rise to paradoxes of higher-order vagueness similar to the standard (first-order) sorites paradox that are not as easily avoided as the basic problem.

According to an emerging consensus on the problem of higher-order vagueness, the culprit is global consequence, which should be relinquished at the expense of its regional counterpart [3]. In particular, Graff Fara's argument employs the argument pattern allowing one to infer *definitely A* from *A*, which fails under a regional consequence relation. However, this argument pattern seems central to the logic of *definitely*. What is more, Elia Zardini [32] has recently shown that the pattern is not needed to derive a higher-order sorites paradox.

To address the problem of higher-order vagueness, we again turn to the multilateral framework. Proof analysis within this framework reveals that the fault lies with some of the rules for the Δ operator. Model-theoretically, this means that the culprit is not global consequence, but the modal properties of Δ. By removing the problematic rules, we obtain a logic in which Δ is non-normal—that is, it does not satisfy $\Delta(A \rightarrow B) \rightarrow (\Delta A \rightarrow \Delta B)$, the modal logic axiom **K**. In particular, the asserted fragment of the resulting logic coincides with the global consequence relation of the non-normal modal logic **NT**. The non-normality of Δ will turn out to be a crucial ingredient in avoiding Zardini's higher-order sorites paradox.

The modal logic **NT** may appear to be too weak to represent the supervaluationist canons of good deductive reasoning. However, its global consequence relation is strong enough to account for the validity of several argument patterns that are beyond reach of its local consequence relation. For instance, although Axiom **K** fails in **NT**, the global definition of validity means that the assertion of ΔB can be inferred from the assertion of ΔA and the assertion of $\Delta(A \rightarrow B)$. Similarly, although $A \rightarrow \Delta A$ fails, the assertion of ΔA can be inferred from the assertion of A, thus vindicating a central argument pattern of the logic of *definitely* which fails under a regional consequence relation.

The difference between an argument and the corresponding material conditional will be a recurring theme of our discussion. In many cases, supervaluationism validates the argument pattern but not the material conditional. As we shall see, the supertruth of a material conditional corresponds to the relevant argument pattern being *locally* valid. Supervaluationists should therefore stress the primacy of global validity *qua* preservation of truth *simpliciter*. The demand to validate material conditionals that correspond to valid argument patterns may stem from the prevalence of standard modal logic, in which truth *simpliciter* is truth at a world and consequence is therefore local.

Proof-theoretically, the failure of principles such as Axiom **K** and $A \rightarrow \Delta A$ and the validity of the corresponding inferences can coexist thanks to the failure of the classical meta-inferences, in particular conditional proof. Supervaluationists should therefore embrace this failure as an integral part of their approach to vagueness, rather than excuse it as an embarrassing but negligible outcome thereof.

The plan is as follows. We begin by explaining the proof-theoretic problem for supervaluationism, paying special attention to the role of the problematic meta-inferences (Section 2). We then present *supervaluationist multilateral logic* and show how it addresses the proof-theoretic problem (Section 3). We go on to introduce the problem of higher-order vagueness (Section 4) and note that the basic version of the problem already forces us to revise supervaluationist multilateral logic. We prove that the resulting logic is sound and complete with respect to **NT** global consequence (Section 5). We explain how these revisions put the supervaluationist in a position to account for vagueness in the same way at all orders (Section 6). We formally execute this plan by extending our logic to the first-order domain. The result is a system which naturally blocks the derivation of Zardini's higher-order sorites paradox (Section 7).

We conclude by commenting on the different roles played by the proof theory and the model theory in our approach (Section 8).

2 Supervaluationism and the Proof-Theoretic Problem

According to a standard way of presenting it, supervaluationism takes vagueness to result from semantic indecision [9, 18]. The use of vague terms is compatible with a range of possible semantic values being assigned to them. The various ways of assigning semantic values to vague terms correspond to different way of making vague terms precise, called *precisifications*. The semantic value of sentences is then determined by quantification over all precisifications. In particular, say that a sentence is *supertrue* if it is true on all precisifications, and *superfalse* if it is false on all precisifications. Then, according to supervaluationism, a sentence is *true simpliciter* if it is supertrue, *false simpliciter* if it is superfalse, and *borderline* if it is neither supertrue nor superfalse.

It is natural for supervaluationists to want to talk about supertruth in the object language. To this end, they are wont to employ an operator Δ, whose intended reading is 'it is definitely the case that'. Roughly, ΔA is true on a precisification just in case A is supertrue. Thus, supervaluationists can say that A is supertrue by stating that it is definitely the case that A. Similarly, they can say that A is superfalse by stating that it is definitely the case that not A. And they can say in the object language that A is borderline by stating that is neither definitely the case that A nor definitely the case that not A.

These ideas can be implemented model-theoretically. We work with a language $\mathcal{L}_\Delta$, obtained by adding Δ to the language of propositional logic. A model is an ordered pair $\langle \mathcal{W}, \mathcal{V} \rangle$, where $\mathcal{W}$ is a non-empty set of points and $\mathcal{V}$ is a valuation function. The points in the model represent precisifications, instead of representing possible worlds as in standard modal logic. The valuation function maps each point to a set of atomic sentences, namely the sentences true at the point. The clauses for the connectives are as usual: $\neg A$ is true at a point w if and only if A is not true at w; $A \wedge B$ is true at w if and only if A is true at w and B is true at w; $A \vee B$ is defined as $\neg(\neg A \wedge \neg B)$, and $A \to B$ is defined as $\neg(A \wedge \neg B)$. The operator Δ serves to talk about supertruth (that is, truth on all precisifications) in the object language. The standard clause for Δ attempts to capture this idea by taking ΔA to be true at a point just in case A is true at all points.[1]

This simple supervaluationist model theory allows us to recursively determine the truth-value of sentences within a given model. However, we are also interested in determining the status of *arguments* within the supervaluationist framework. We focus on arguments with one or more premisses and a single conclusion but will indicate when important differences would arise should multiple-conclusion arguments

[1] We will see below that this clause for Δ is problematic given the phenomenon of higher-order vagueness. We will argue that the clause can be revised to avoid the problem of higher-order vagueness while maintaining Δ's role as an object language representation of supertruth.

be countenanced.[2] In the context of modal logic, it is customary to define validity in terms of preservation of truth at a point. The resulting notion is known as *local validity* and is formally defined as follows.

Definition 1 (Local validity) Let Γ be a set of $\mathcal{L}_{\Delta}$-sentences and A an $\mathcal{L}_{\Delta}$-sentence. Then the argument from Γ to A is *locally valid* (written $\Gamma \models_l A$) iff for every model $\langle \mathcal{W}, \mathcal{V} \rangle$ and $w \in \mathcal{W}$, if C is true at w on $\mathcal{V}$ for every $C \in \Gamma$, then A is true at w on $\mathcal{V}$.

The reason for the adoption of local validity in modal logic is that truth *simpliciter* is identified with truth at a world, and validity is taken to consist in preservation of truth *simpliciter*. In the context of supervaluationism, however, truth *simpliciter* is identified with supertruth (that is, truth on all precisifications), which suggests adopting a *global* definition of validity instead.

Definition 2 (Global validity) Let Γ be a set of $\mathcal{L}_{\Delta}$-sentences and A an $\mathcal{L}_{\Delta}$-sentence. Then the argument from Γ to A is *globally valid* (written $\Gamma \models_g A$) iff for every model $\langle \mathcal{W}, \mathcal{V} \rangle$, if C is true at w on $\mathcal{V}$ for every $C \in \Gamma$ and every $w \in \mathcal{W}$, then A is true at w on $\mathcal{V}$ for every $w \in \mathcal{W}$.

Global validity allows the supervaluationist to hold on to the idea that validity consists in the preservation of truth *simpliciter*. Global validity and local validity coincide for the Δ-free fragment of the language.[3] Moreover, every locally valid argument is also globally valid in the entire supervaluationist language.[4] However, there are globally valid arguments that are not locally valid. In particular, the argument

$$A \therefore \Delta A \qquad (\Delta\text{-Strengthening})$$

is globally but not locally valid. The global validity of Δ-Strengthening can be exploited to produce counterexamples to certain classical *meta-inferences* involving dischargeable assumptions.

Conditional proof	$A \models_g \Delta A$, but $\not\models_g A \rightarrow \Delta A$
Reductio	$A \wedge \neg\Delta A \models_g \Delta A$ and $A \wedge \neg\Delta A \models_g \neg\Delta A$, but $\not\models_g \neg(A \wedge \neg\Delta A)$
Proof by cases	$A \models_g \Delta A \vee \Delta\neg A$ and $\neg A \models_g \Delta A \vee \Delta\neg A$, but $A \vee \neg A \not\models_g \Delta A \vee \Delta\neg A$
Contraposition	$A \models_g \Delta A$, but $\neg\Delta A \not\models_g \neg A$

[2] The existence of multiple-conclusion arguments in the wild is controversial, as are the credentials of multiple-conclusion systems as generalisations of actual inferential practice. For discussion, see [7, 23].

[3] The focus on single-conclusion arguments is crucial here. For instance, let A be an atomic borderline sentence. Then the multiple-conclusion Δ-free argument from $A \vee \neg A$ to $\{A, \neg A\}$ is locally but not globally valid.

[4] For if an argument is not globally valid, then there is a model in which the premisses are true on all precisifications and the conclusion is false on some precisification w. In such a model, there is *a fortiori* a precisification in which the premisses are true and the conclusion false, namely w.

In a first-order setting, there are counterexamples to the meta-inference of existential instantiation [28].

Existential instantiation $Fa \wedge \neg \Delta Fa \models_g \bot$, but $\exists x(Fx \wedge \neg \Delta Fx) \not\models_g \bot$

In fact, the meta-inferences *must* fail if global consequence is not to collapse into local consequence. For suppose that $A \models_g B$. If conditional proof were globally valid, we could conclude $\models_g A \rightarrow B$. But this means that, for every model and every point w in the model, if A is true at w, then so is B. That is, $A \models_l B$. Conversely, $A \models_l B$ immediately entails $A \models_g B$. Thus conditional proof must fail if global consequence is not to reduce to local consequence.[5] Similar considerations apply to the other meta-inferences. The failure of the meta-inferences is part and parcel of the supervaluationist approach based on global consequence.

We have considered validity as preservation of truth at a point, and validity as preservation of truth at all points. But the analogy with the case of modal logic suggests a third possibility, namely validity as preservation of truth at all *accessible* points. As usual, which points are accessible is determined by an accessibility relation R and models are therefore triples $\langle \mathcal{W}, \mathcal{V}, R \rangle$. The resulting notion of *regional validity* can be formally defined thus.

Definition 3 (Regional validity) Let Γ be a set of $\mathcal{L}_\Delta$-sentences and A an $\mathcal{L}_\Delta$-sentence. Then the argument from Γ to A is *regionally valid* (written $\Gamma \models_r A$) iff for every model $\langle \mathcal{W}, \mathcal{V}, R \rangle$ and $w \in \mathcal{W}$, if C is true at v on $\mathcal{V}$ for every $C \in \Gamma$ and every $v \in \mathcal{W}$ such that wRv, then A is true at v on $\mathcal{V}$ for every $v \in \mathcal{W}$ such that wRv.

Validity may still be identified with preservation of supertruth, but supertruth is modeled as truth at all accessible points. Points can now disagree about which propositions are supertrue. Given that the Δ-operator serves to express supertruth in the object language, points should also disagree about what they take to be definitely the case. The definition of Δ is modified accordingly, and ΔA is true at a point w iff A is true at every v such that wRv.

If we only require the accessibility relation R to be reflexive, Δ-Strengthening is not regionally valid [3, p. 305]. But one can still construct counterexamples to the aforementioned classical meta-inferences using the regional validity of

$$A \wedge \neg \Delta A \therefore \bot \qquad (\Delta\text{-Contradiction})$$

The reader is referred to [3, 28] for details. Given that adopting regional validity does not affect the failure of classical meta-inferences, we will begin by focusing on global validity, which has often been considered the correct notion of validity for the

[5] This is not to say that the supervaluationist must deny that the indicative conditional satisfies conditional proof, since the natural language conditional need not be the material conditional. Of course, the supervaluationist should then provide an account of the indicative conditional which allows her to hold on to conditional proof for the indicative whilst avoiding the collapse to local consequence. We explore the prospects for such an account in ongoing work.

supervaluationist to adopt; see, e.g. [9, 16]. We shall return to regional validity when discussing the problem of higher-order vagueness in Section 4.

The failure of classical meta-inferences has been a source of concerns about supervaluationism. One concern has been that the failure of classical meta-inferences makes supervaluationism revisionary of classical logic. J. Robert G. Williams [26] has argued that supervaluationists can avoid the failures of the classical meta-inferences by treating Δ as a non-logical constant, that is by allowing its meaning to vary across interpretations. As Nicholas Jones [15, pp. 638–639] points out, however, this move prevents supervaluationists from investigating argument patterns specific to vague discourse (such as Δ-Strengthening and Δ-Contradiction), which should be one of the aims of supervaluationist logic. Instead, supervaluationists should simply insist that to properly account for these argument patterns, the classical meta-inferences must be revised.

The more pressing concern is whether the failure of classical meta-inferences renders supervaluationism unable to provide an account of good deductive reasoning [10, 28]. To address this concern, supervaluationists must specify which argument patterns govern good deductive reasoning when vague terms are involved, if these are not the classical ones. Formally, this amounts to providing a solution to what we call the *proof-theoretic problem* for supervaluationism. This is the problem of providing a proof system which is sound and complete with respect to supervaluationist model theory and which includes supervaluationistically acceptable versions of the problematic meta-inferences. To count as a solution to the proof-theoretic problem, the proof system must codify canons of reasoning which are cognitively feasible to use in actual deductive practice. Ideally, the proof system should consist of rules satisfying standard proof-theoretic constraints on their admissibility, notably harmony (see [8]).

A natural strategy to tackle the proof-theoretic problem would seem to take a standard natural deduction system for classical logic and restrict its meta-inferences to the Δ-free language. To account for deductive reasoning involving *definitely*, one can then add supervaluationistically acceptable versions of the problematic meta-inferences, holding for the entire supervaluationist language $\mathcal{L}_\Delta$. Rosanna Keefe [16, pp. 179–181] has suggested that suitable versions of the meta-inferences may be obtained by inserting Δ at appropriate places.

$$(\to\text{I}^*.)\ \frac{\begin{matrix}[A]\\ \vdots\\ B\end{matrix}}{\Delta A \to B} \qquad\qquad (\neg\text{I}^*.)\ \frac{\begin{matrix}[A]\\ \vdots\\ \bot\end{matrix}}{\neg\Delta A}$$

$$(\vee\text{E}^*.)\ \frac{\Delta A \vee \Delta B \quad \begin{matrix}[A]\\ \vdots\\ C\end{matrix} \quad \begin{matrix}[B]\\ \vdots\\ C\end{matrix}}{C} \qquad\qquad (\text{Contraposition}^*)\ \frac{\begin{matrix}[A]\\ \vdots\\ B\end{matrix} \quad \neg B}{\neg\Delta A}$$

Even in the presence of Axiom **K** for Δ, however, the resulting natural deduction system is incomplete with respect to supervaluationist model theory under a global notion of validity. For instance, in this system $\Delta q \vee p$ is not derivable from $p \vee \Delta q$, despite this argument being globally valid. Of course, this argument is a substitution instance of a classically valid argument, but this is just a confirmation of the fact that we have not succeeded in providing a proof theory that axiomatises the global consequence relation. The problem is that in order to derive $p \vee \Delta q$ from $\Delta q \vee p$ one would like to use classical proof by cases, which is only available in the Δ-free language.

The reason why, in this case, the relevant application of classical proof by cases should count as acceptable even by supervaluationist lights is that it rests on hypothetical reasoning which does not employ distinctively supervaluationist inferential moves.[6] This suggests that to axiomatise global consequence the supervaluationist should place restrictions not on the language to which the classical meta-inferences apply, but on the inferential moves that are allowed within their subderivations.

An approach based on restricting inferential moves (but not specifically within hypothetical proof contexts) is recommended by Pablo Cobreros [4]. He shows that by adding Δ-Strengthening as a rule to a proof system that is sound and complete for local consequence, one obtains a system complete for global consequence. To ensure that this system is also sound for global consequence, Cobreros continues, one must then restrict the locally but not globally valid rules of inference so that they can only be applied if Δ-Strengthening has not been previously applied in the derivation.[7] We have independently developed an approach based on restricting inferential moves to axiomatise global consequence in the context of the study of epistemic *might* [13], where global consequence is known as *informational consequence* [30]. In our earlier work, we provide a natural deduction system for epistemic modality in which only one primitive rule must be restricted and the restrictions are confined to hypothetical proof contexts. The system has the distinctive feature of also satisfying standard proof-theoretic criteria on the acceptability of inference rules. In the next section, we show how this strategy can be used to provide a proof-theoretically satisfying axiomatisation of global consequence in the context of supervaluationism.

3 Supervaluationist Multilateral Logic

Let p stand for a simple sentence involving a vague predicate, such as *Harry is bald*. What basic attitudes may one hold towards p? One may believe that p and hence *assert* p. One may disbelieve that p and hence assert its *negation*. But one may refrain from believing that p and from disbelieving that p. Accordingly, one may be willing to assert neither p nor its negation.

We can easily make sense of these attitudes in supervaluationist terms. According to the supervaluationist, one may believe that p is (super)true and hence assert p. One may believe that p is super(false) and hence assert its negation. But one may

[6] Williams [26, §7] makes a similar observation from the model-theoretic perspective.

[7] Cobreros and Tranchini [6] extend Cobreros's strategy to cut-free, multiple conclusion calculi.

believe that p is borderline and hence refrain from believing that p and refrain from disbelieving that p. Hence, one may be willing to assert neither p nor its negation.

We call *rejection* the speech act whereby one expresses refraining from believing that A (see our [12]) and *weak assertion* the speech act whereby one expresses refraining from disbelieving that A (see our [13]). The supervaluationist can express her state of mind when presented with a borderline sentence by rejecting and weakly asserting that sentence.

We can use these ideas to develop a deductive system for supervaluationist logic. We work with a language $\mathcal{L}^S_\Delta$ in which formulae are *signed*, that is prefixed by signs for assertion, rejection or weak assertion. The language is defined as follows. We say that A is a *sentence* of $\mathcal{L}^S_\Delta$ if it is a member of the unsigned language $\mathcal{L}_\Delta$. And we say that φ is a *formula* of $\mathcal{L}^S_\Delta$ if it is either $\perp$ or the result of prefixing a $\mathcal{L}^S_\Delta$-sentence with one of $+, \ominus, \oplus$. These are *force-markers* standing, respectively, for assertion, rejection and weak assertion. The absurdity sign $\perp$ is treated as neither a sentence nor a (0-place) connective, but as a punctuation mark indicating that a logical dead end has been reached [20, 24]. Throughout, uppercase Latin letters denote sentences and lowercase Greek letters denote formulae.

Our natural deduction system is in the style of Gentzen. *Bilateral* systems [12, 20, 22] use rules that specify conditions on assertion and rejection.[8] Our deductive system specifies conditions on assertion, rejection and weak assertion and is therefore *multilateral* (see [13]). The rules we shall lay down for the logical constants satisfy the standard proof-theoretic constraints on their admissibility, *viz.* purity, simplicity and harmony (see [8, 20]). One may therefore regard these rules as giving the meaning of the constants featuring in them, as opposed to this meaning being given by the relevant model-theoretic clauses.[9]

We begin with the connectives. The rules for conjunction are the usual ones except that sentences are now prefixed by the assertion sign.

$$(+\wedge\text{I.})\ \frac{+A \qquad +B}{+A \wedge B}\quad (+\wedge\text{E.}_1)\ \frac{+A \wedge B}{+A}\quad (+\wedge\text{E.}_2)\ \frac{+A \wedge B}{+B}$$

Next we have rules specifying how to introduce and eliminate negations by moving from weak assertion to rejection, and from rejection to weak assertion. The rules ensure that one ought to refrain from believing A just in case one ought to refrain from disbelieving its negation, and that one ought to refrain from disbelieving A just in case one ought to refrain from believing its negation.

$$(\ominus\neg\text{I.})\ \frac{\oplus A}{\ominus\neg A}\qquad (\ominus\neg\text{E.})\ \frac{\ominus\neg A}{\oplus A}\qquad (\oplus\neg\text{I.})\ \frac{\ominus A}{\oplus\neg A}\qquad (\oplus\neg\text{E.})\ \frac{\oplus\neg A}{\ominus A}$$

This makes supervaluationist sense. If one holds a sentence not to be false, then one ought to hold its negation not to be true, and *vice versa*. Similarly, if one holds a sentence not to be true, then one ought to hold its negation not to be false, and *vice versa*.

[8] Bilateral systems are premised on the idea that rejection cannot be identified with negative assertion. This is compatible with the Equivalence Thesis that a rejection can be inferred from the corresponding negative assertion and *vice versa*. Smiley's [22] and Rumfitt's [20] bilateral systems satisfy the Equivalence Thesis, whereas ours [12] does not, in that it allows for rejections which do not imply the corresponding negative assertion. The systems developed in this paper follow our previous work.

[9] We discuss the relation between multilateral logic and proof-theoretic constraints at some length in [13].

The definitely operator serves to express supertruth in the object language. Accordingly, we have rules that allow one to pass from the assertion of A to the assertion of *definitely* A and *vice versa*.

$$(+\Delta\text{I.})\ \frac{+A}{+\Delta A} \qquad (+\Delta\text{E.})\ \frac{+\Delta A}{+A}$$

Thus, our logic immediately sanctions Δ-Strengthening—as it should, given its global validity. However, defenders of regional validity have challenged the supervaluationist acceptability of Δ-Strengthening, and so we shall discuss its status in the next section.

The rules for asserted Δ fail to capture the full strength of its standard model-theoretic clause. Recall that, when working with a global definition of validity, Δ is typically defined so that ΔA is true at a point just in case A is true at all points. According to this definition, someone ought to hold A to be true not only if they hold ΔA to be true, but also if they simply hold ΔA not to be false. This means that one should be able to infer the assertion of A not only from the assertion of ΔA, but also from its weak assertion. We may ensure this is the case with a suitable elimination rule for weak assertions of *definitely* A, together with a corresponding introduction rule.

$$(\oplus\Delta\text{I.})\ \frac{+A}{\oplus\Delta A} \qquad (\oplus\Delta\text{E.})\ \frac{\oplus\Delta A}{+A}$$

We are not quite done yet. As is customary in multilateral systems, we also need to lay down *coordination principles* for the relevant speech acts. The first set of coordination principles ensures that assertion and rejection are contradictories.

$$(\text{Rejection})\ \frac{+A \qquad \ominus A}{\bot} \quad (\text{SR}_1)\ \frac{\begin{matrix}[+A]\\ \vdots\\ \bot\end{matrix}}{\ominus A} \quad (\text{SR}_2)\ \frac{\begin{matrix}[\ominus A]\\ \vdots\\ \bot\end{matrix}}{+A}$$

(Rejection) states that it is absurd to both believe and refrain from believing that A. (SR_1) states that if one's believing that A leads to absurdity, then one ought to refrain from believing that A. Similarly for (SR_2).

The second set of coordination principles ensures that weak assertion is subaltern to assertion *simpliciter*. We write $\overset{+}{\vdots}$ for a derivation in which all premisses and undischarged assumptions are of the form $+A$. Since $\bot$ is treated as a punctuation mark, we distinguish in (Weak Inference) between the case in which the subderivation leads to $+B$ and the case in which it leads to a logical dead end. In the former case, one is allowed to conclude $\oplus B$ given $\oplus A$; in the latter case, one is allowed to conclude that a logical dead end has been reached, again given $\oplus A$.

$$(\text{Assertion})\ \frac{+A}{\oplus A} \quad (\text{Weak Inference})\ \frac{\oplus A \qquad \begin{matrix}[+A]\\ \overset{+}{\vdots}\\ +B/\bot\end{matrix}}{\oplus B/\bot} \ \text{if } (+\Delta\text{I.}) \text{ and } (\oplus\Delta\text{I.}) \text{ were not used to derive } +B.$$

(Assertion) states that if one believes A, then one ought to refrain from disbelieving A. (Weak Inference) states that if one refrains from disbelieving A, then one ought to hold the same attitude towards all of A's consequences. Note that (Weak Inference) disallows the use of the introduction rules for Δ in its subderivation. We will shortly make use of this fact to provide supervaluationistically acceptable versions of the problematic meta-inferences.

This concludes the exposition of the rules of our logic. That is, we let *Supervaluationist Multilateral Logic* (SML for short) be the natural deduction system consisting of the introduction and elimination rules for $\wedge$, $\neg$ and Δ and the coordination principles for assertion, rejection and weak assertion.

SML is sound and complete with respect to **S5** *modulo* a suitable translation. In particular, let τ be the following mapping from $\mathcal{L}^S_\Delta$-formulae to $\mathcal{L}_\Delta$-formulae

$$\tau(\varphi) = \begin{cases} \Delta\psi, \text{ if } \varphi = +\psi \\ \neg\Delta\psi, \text{ if } \varphi = \ominus\psi \\ \neg\Delta\neg\psi, \text{ if } \varphi = \oplus\psi. \end{cases}$$

Moreover, write $\tau[\Gamma]$ for $\{\tau(\varphi) : \varphi \in \Gamma\}$ and let $\vdash_{\mathsf{SML}}$ denote derivability in SML. We have:

Theorem 1 *For any $\mathcal{L}^S_\Delta$-formula φ and set of $\mathcal{L}^S_\Delta$-formulae Γ, $\Gamma \vdash_{\mathsf{SML}} \varphi$ iff $\tau[\Gamma] \models_l \tau(\varphi)$.*

The proof is a straightforward adaption of the proof we give in [14], where we show that EML—the system obtained by replacing the Δ rules with the following rules for $\Diamond$—is sound and complete with respect to **S5** *modulo* a suitable translation.

$$(+\Diamond\text{I.})\ \frac{\oplus A}{+\Diamond A} \qquad (+\Diamond\text{E.})\ \frac{+\Diamond A}{\oplus A} \qquad (\oplus\Diamond\text{I.})\ \frac{\oplus A}{\oplus\Diamond A} \qquad (\oplus\Diamond\text{E.})\ \frac{\oplus\Diamond A}{\oplus A}$$

Theorem 1 immediately yields:

Proposition 1 *For any $\mathcal{L}_\Delta$-formula A and set of $\mathcal{L}_\Delta$-formulae Γ, $\{+B \mid B \in \Gamma\} \vdash_{\mathsf{SML}} +A$ iff $\Gamma \models_g A$.*

The SML logic of assertion is supervaluationist logic under a global notion of validity.

What is the status of the supervaluationistically problematic meta-inferences in SML? Earlier we made the claim that the supervaluationist can provide supervaluationistically acceptable versions of the problematic meta-inferences by appropriately restricting the application of certain distinctive rules. We can now make good on that claim: the restrictions on (Weak Inference) carry over to the problematic meta-inferences.

Proposition 2 *The following rules are derivable in* SML.

$$(+\rightarrow\text{I.})\ \frac{\begin{matrix}[+A]\\ +\vdots \\ +B\end{matrix}}{+A \rightarrow B}\ \text{if } (+\Delta\text{I.}) \text{ and } (\oplus\Delta\text{I.}) \text{ were not used to derive } +B \qquad (+\neg\text{I.})\ \frac{\begin{matrix}[+A]\\ +\vdots \\ \bot\end{matrix}}{+\neg A}\ \text{if } (+\Delta\text{I.}) \text{ and } (\oplus\Delta\text{I.}) \text{ were not used to derive } \bot$$

$$(+\vee\text{E.})\ \frac{+A\vee B \qquad \begin{matrix}[+A]\\ +\vdots \\ +C\end{matrix} \qquad \begin{matrix}[+B]\\ +\vdots\\ +C\end{matrix}}{+C}\ \text{if } (+\Delta\text{I.}) \text{ and } (\oplus\Delta\text{I.}) \text{ were not used to derive } +C$$

$$(+\text{Contraposition})\ \frac{\begin{matrix}[+A]\\ +\vdots \\ +B\end{matrix} \qquad +\neg B}{+\neg A}\ \text{if } (+\Delta\text{I.}) \text{ and } (\oplus\Delta\text{I.}) \text{ were not used to derive } +B$$

The supervaluationistically acceptable versions of the meta-inferences are derivable from the rules characterizing the meaning of the logical operators together with the coordination principles. The technique is similar in each case. We present here the derivation of (+¬I.).

$$\text{where } (+\Delta\text{I.}) \text{ and } (\oplus\Delta\text{I.}) \text{ were not used to derive } \bot \qquad \cfrac{\cfrac{\begin{matrix}[+A]^1\\ +\vdots\\ \bot\end{matrix} \qquad \cfrac{[\ominus\neg A]^2}{\oplus A}\,(\ominus\neg\text{E.})}{\bot}\,(\text{Weak Inference})^1}{+\neg A}\,(\text{SR}_2)^2$$

Thus, the supervaluationist can provide a proof-theoretically satisfying, sound and complete proof system for her logic which includes supervaluationistically acceptable versions of the meta-inferences. This solves the proof-theoretic problem for supervaluationism.[10]

We have so far ignored what is perhaps the standard objection to supervaluationism, namely its difficulties with higher-order vagueness. We are going to show that far from being a hindrance, the failure of the standard versions of the meta-inferences can help us reconcile supervaluationism with higher-order vagueness. We begin by introducing the problem of higher-order vagueness.

4 The Problem of Higher-Order Vagueness

The predicate *bald* is vague. Thus, the supervaluationist holds, it may be true that Harry is bald, it may be false, but it may also be borderline—it may be neither true nor false that Harry is bald. Using the Δ operator, the supervaluationist can express this fact by saying that we may have ΔA, we may have $\Delta\neg A$ but we may also have $\neg\Delta A \wedge \neg\Delta\neg A$. It is not a theorem that $\Delta A \vee \Delta\neg A$.

The predicate *definitely bald* also appears to be vague. Thus, the supervaluationist ought to hold, it may be true that Harry is definitely bald, it may be false, but it may also be borderline—it may be neither true nor false that Harry is definitely bald. Using the Δ operator, the supervaluationist would like to convey this fact by saying that we may have $\Delta\Delta A$, we may have $\Delta\neg\Delta A$ but we may also have $\neg\Delta\Delta A \wedge$

[10] Note, in addition, that the supervaluationistically acceptable versions of the meta-inferences apply throughout the entire supervaluationist language. Thus, supervaluationists are not forced to postulate two distinct modes of reasoning, one for the Δ-free language and another for reasoning involving Δ. This circumvents worries raised by Varzi [25, p. 657] and Williamson [28, pp. 535–536] among others.

$\neg\Delta\neg\Delta A$. However, this is precluded by the standard model-theoretic clause for Δ, which entails that $\Delta\Delta A \vee \Delta\neg\Delta A$ is globally (as well as locally) valid. It is a theorem that $\Delta\Delta A \vee \Delta\neg\Delta A$.

Supervaluationism seemingly leaves no room for indefinite cases of definite baldness. This is the *basic* problem of higher-order vagueness. To address it, a natural strategy for the supervaluationist is to adopt the regional definition of Δ, according to which ΔA is true at a point just in case it is true at every point accessible from w. If the accessibility relation is non-transitive or non-symmetric, $\Delta\Delta A \vee \Delta\neg\Delta A$ is no longer a theorem. Within the context of standard modal logic, this means that one of the modal logic axioms **4** and **B** must be abandoned.

However, an argument due to Graff Fara [10] shows that the supervaluationist's difficulties with higher-order vagueness do not end here. Consider a sorites series involving *bald*. At the beginning of the series, we have a_1, who is bald. At the end of the series, we have a_m, who isn't bald. The transition from beginning to end appears not to be *sharp*. The supervaluationist aims to explain this apparent lack of a sharp boundary with the presence of truth-value gaps. Accordingly, Graff Fara argues, the supervaluationist is committed to the principle that if someone is definitely bald, then their successor in the series is not definitely not bald. Letting F stand for the predicate *bald*, we can formalise the principle thus.

$$\Delta Fa_j \rightarrow \neg\Delta\neg Fa_{j+1} \qquad (F\text{-Gap})$$

What is more, the entire transition from baldness to not baldness appears not to be sharp. Vagueness seems to cut at every order. Thus, says Graff Fara, the supervaluationist is also committed to a gap principle for ΔF and indeed for any finite iteration of Δ.

$$\Delta\Delta^n Fa_j \rightarrow \neg\Delta\neg\Delta^n Fa_{j+1} \qquad (\text{Gap})$$

Since $\models_g \varphi \rightarrow \neg\psi$ just in case $\models_g \psi \rightarrow \neg\varphi$, the gap principles are equivalent to the following principle.

$$\Delta\neg\Delta^n Fa_{j+1} \rightarrow \neg\Delta\Delta^n Fa_j \qquad (\text{Contraposed Gap})$$

However, the gap principles are inconsistent with Δ-Strengthening. Take the first member of the sorites series, a_1. It is true that a_1 is bald, that is Fa_1. By applying Δ-Strengthening to Fa_1 m-many times we obtain $\Delta^m Fa_1$. Now take the end member of the sorites series, a_m. It is false that a_m is bald, that is $\neg Fa_m$. By Δ-Strengthening, it follows that $\Delta\neg Fa_m$. Together with the (Contraposed) Gap Principle for F, this implies that $\neg\Delta Fa_{m-1}$. By Δ-Strengthening, it follows that $\Delta\neg\Delta Fa_{m-1}$. Together with the Gap Principle for Δ, this implies that $\neg\Delta^2 Fa_{m-2}$. By repeated applications of Δ-Strengthening and the relevant gap principle, we eventually obtain $\neg\Delta^m Fa_1$, contradicting $\Delta^m Fa_1$. Even if $\Delta\Delta A \vee \Delta\neg\Delta A$ is not a theorem, supervaluationism appears to be inconsistent with unrestricted higher-order vagueness.

Graff Fara's argument makes use of Δ-Strengthening, which, as mentioned above, is not regionally valid when the accessibility relation is only reflexive. Regional validity is naturally coupled with a regional definition of Δ, which, given the same constraints on the accessibility relation, means that $\Delta\Delta A \vee \Delta\neg\Delta A$ is no longer a theorem. For this reason, some have suggested that the supervaluationist should account for higher-order vagueness by adopting a regional characterisation of validity (see, e.g., [3, 28]).

Not validating Δ-Strengthening is a high price to pay. Intuitively, if one is willing to assert that A is the case, then one ought to be willing to assert that A is definitely the case. Putting things in more theoretical terms, the function of the Δ operator is to enable the supervaluationist to say in the object language that A is supertrue, by asserting ΔA. As Graff Fara [10, pp. 199–200] has pointed out, this means that Δ-Strengthening should have for the supervaluationist the same status as the T-Intro rule, and so should not be jettisoned lightly.

Cobreros offers two considerations to make the failure of Δ-Strengthening more palatable [3, pp. 306–307]. The first consideration is that we can explain away the intuitive appeal of Δ-Strengthening. In particular, says Cobreros, our intuitions come from a mistaken use of *reductio* (which is not regionally valid) to derive Δ-Strengthening from the fact that A and $\neg\Delta A$ are regionally inconsistent. The defender of global validity is likely to reply that the reason why we think that A and $\neg\Delta A$ are inconsistent in the first place is that we find Δ-Strengthening intuitively valid. Be that as it may, we take it that the defender of regional validity should provide some evidence that ordinary speakers do use *reductio* where they should not. It is not enough to simply postulate that speakers make such mistakes to explain away the intuitive appeal of Δ-Strengthening.

Cobreros's second consideration is that the two ingredients of the model theory that result in the invalidity of Δ-Strengthening—namely, the identification of validity with regional validity and the non-transitivity of the accessibility relation—are well motivated by the phenomenon of higher-order vagueness. In particular, says Cobreros, the phenomenon demonstrates that whether a sentence is true is itself vague, and this vagueness must reside in the fact that what counts as an admissible precisification is vague. This, in turn, must be captured by taking supertruth to be truth at all *accessible* points. Moreover, Cobreros continues, the accessibility relation must not be transitive to make room for higher-order vagueness.

Cobreros is right that to account for higher-order vagueness, the supervaluationist must be able to make room for the truth of a sentence being a vague matter. However, given that truth in the object language is expressed by the Δ operator, this simply means that $\neg\Delta\Delta A \wedge \neg\Delta\neg\Delta A$ must be consistent. And the consistency of $\neg\Delta\Delta A \wedge \neg\Delta\neg\Delta A$ can be achieved by taking Δ to be regionally defined and the accessibility relation to be only reflexive. It does not require that validity be regionally defined as well: a regional definition of Δ is compatible with truth *simpliciter* being global, i.e. being identified with truth at all points. One may object that defining Δ to be regional and truth *simpliciter* to be global means that Δ is not the object language representation of truth. But this is not so: provided the accessibility relation is reflexive, a regionally defined Δ operator is still an object language representation of truth *simpliciter* because the supertruth of ΔA implies the supertruth of A, and *vice versa*.

Thus, it is possible to (i) hold on to a global notion of validity, (ii) adopt a regional definition of Δ to deal with the phenomenon of higher-order vagueness and (iii) respect the insight that Δ is an object language representation of truth. The reason is that what matters to determine whether Δ is an object language representation of truth is not its local behaviour, but its global one, since truth *simpliciter* is identified with truth at all points, rather than with truth at a point as in standard modal logic. For the supervaluationist, the local definition of an expression matters only insofar as it serves to determine its global behaviour. And while the local definition of Δ does not match the definition of truth *simpliciter*, the regionally defined Δ behaves globally like truth (as long as its accessibility relation is reflexive). One advantage of our proof-theoretic approach is that it sidesteps the need to employ local definitions at all: the clauses of our proof theory only concern the global behaviour of the relevant expressions.

Let us take stock. Supervaluationism must invalidate $\Delta\Delta A \vee \Delta\neg\Delta A$ because of basic higher-order vagueness. This may be ensured by taking the accessibility relation in the definition of Δ to be non-transitive or non-symmetric. But the accessibility relation may still be reflexive. Since this regional definition of Δ is compatible with consequence being global and with Δ being an object language representation of truth, the basic problem of higher-order vagueness does not force the rejection of Δ-Strengthening upon us. But can the validity of Δ-Strengthening be reconciled with genuine higher-order vagueness? We now take up this task.

5 Meta-Inferences and Basic Higher-Order Vagueness

The principle that if one is willing to assert that A is the case, then one ought to be willing to assert that A is definitely the case withstood initial scrutiny. The principle sanctions the (+ΔI.) rule, and our task is now to show how this rule can be reconciled with higher-order vagueness.

The basic problem of higher-order vagueness already requires $+\neg\Delta\Delta A \wedge \neg\Delta\neg\Delta A$ to be consistent. But SML proves $+\Delta\Delta A \vee \Delta\neg\Delta A$. Proof analysis reveals that the argument makes essential use of the elimination rule for weak assertions of *definitely* A, which we repeat here together with the corresponding introduction rule.

$$(\oplus\Delta\text{I.})\ \frac{+A}{\oplus\Delta A} \qquad (\oplus\Delta\text{E.})\ \frac{\oplus\Delta A}{+A}$$

When we introduced the rule in Section 3, we motivated it simply on the basis of the standard model-theoretic clause for Δ. According to this clause, the meaning of ΔA is A *is (super)true*: ΔA is true at a point just in case A is true at all points. The ($\oplus\Delta$E.) rule, in particular, encapsulates the left-to-right direction of the model-theoretic clause in allowing one to move from the *local* claim that ΔA is true at a point to the *global* claim that A is true at all points.

But this is precisely what should fail if we want to make room for higher-order vagueness and hold on to the (+ΔI.) rule. For if one holds ΔA to be borderline, one ought to hold ΔA not to be false. But it should not follow from that one ought to hold A to be true. For otherwise, by the (+ΔI.) rule, one ought to hold ΔA to be true too, which is incompatible with holding ΔA not to be true and hence with holding it to be borderline. Formally, if ΔA is borderline we should have that $+\neg\Delta\Delta A \wedge \neg\Delta\neg\Delta A$. But $+\neg\Delta\neg\Delta A$ requires that somewhere it should be locally the case that ΔA, that is $\oplus\Delta A$. It should not follow from this that A is globally the case, that is $+A$. For otherwise, by the $(+\Delta\mathrm{I.})$ rule, we have that $+\Delta A$, which contradicts ΔA's borderline status and in particular $+\neg\Delta\Delta A$.

We must forfeit the $(\oplus\Delta)$ rules.[11] This addresses the basic problem of higher-order vagueness. Nonetheless, the resulting logic SML^{-} validates Δ-Strengthening, since it includes the $(+\Delta\mathrm{I.})$ rule. This means in particular that $+\Delta A \vdash_{\mathsf{SML}^{-}} +\Delta\Delta A$. But because of the restrictions on conditional proof in SML, this does not entail $\vdash_{\mathsf{SML}^{-}} +\Delta A \rightarrow \Delta\Delta A$, the modal logic Axiom **4**, which would allow to derive $+\Delta\Delta A \vee \Delta\neg\Delta A$. To establish that there is no other derivation of this latter formula, we now develop a model theory for SML^{-}.

The model theory will show that when the meaning of Δ is given by $(+\Delta\mathrm{I.})$ and $(+\Delta\mathrm{E.})$ alone, then Δ is locally an **NT** modality. **NT** is the non-normal, non-classical modal logic that contains all classical tautologies, contains axiom **T** ($\Delta A \rightarrow A$) and is closed under *modus ponens*, substitution and axiom **N** ($\vdash A$ entails $\vdash \Delta A$). It is non-normal because it does not contain axiom **K** and non-classical because it is not closed under axiom **E** ($\vdash A \leftrightarrow B$ entails $\vdash \Delta A \leftrightarrow \Delta B$). Model-theoretically, every sentence is assigned its own accessibility relation. In normal, classical modal logics, each world is associated with a set of accessible worlds; in non-normal, classical modal logics each world and *proposition* is associated with such a set;[12] and in non-normal, non-classical modal logic, each world and *sentence* is associated with such a set, so the *syntactic* representation matters.

Definition 4 An SML^{-}-model is a triple $\mathcal{M} = \langle \mathcal{W}, \mathcal{V}, R\rangle$ where $\mathcal{W}$ is a non-empty set of points, $\mathcal{V}$ is a valuation function and $R : W \times \mathcal{L}_{\Delta} \rightarrow \mathcal{P}(\mathcal{W})$ is a function mapping every point and sentence to a set of points such that for all $w \in \mathcal{W}$ and A, $w \in R(w, A)$.

The local satisfaction-conditions are the same as in **NT**.

- $\mathcal{M}, w \Vdash p$ iff $p \in \mathcal{V}(w)$.
- Boolean connectives as usual.
- $\mathcal{M}, w \Vdash \Delta A$ iff for all $v \in R(w, A)$, $\mathcal{M}, v \Vdash A$.

In the case of **S5**, global consequence could be reduced to a translation into the local language. In the case of **NT**, this is not possible because of the failure of transitivity. We therefore define the global satisfaction-conditions for the force-markers directly.

- $\mathcal{M} \models\!\models +A$ iff for all $w \in \mathcal{W}$, $\mathcal{M}, w \Vdash A$.
- $\mathcal{M} \models\!\models \oplus A$ iff for some $w \in \mathcal{W}$, $\mathcal{M}, w \Vdash A$.
- $\mathcal{M} \models\!\models \ominus A$ iff for some $w \in \mathcal{W}$, $\mathcal{M}, w \nVdash A$.

SML^{-} is sound and complete with respect to this model theory. Write $\Gamma \models_{\mathsf{SML}^{-}} \varphi$ iff for all SML^{-}-models $\mathcal{M}$ with $\mathcal{M} \models\!\models \psi$ for all $\psi \in \Gamma$, it is the case that $\mathcal{M} \models\!\models \phi$.

[11] The $(\oplus\Delta\mathrm{I.})$ rule can be immediately recovered by applying the $(+\Delta\mathrm{I.})$ rule and (Assertion).

[12] Or, equivalently, every world is associated with a *neighborhood*, a set of sets of worlds, representing the propositions necessary at that world.

Theorem 2 *For any $\mathcal{L}^S_\Delta$-formula φ and set of $\mathcal{L}^S_\Delta$-formulae Γ, $\Gamma \vdash_{\mathsf{SML}^-} \varphi$ iff $\Gamma \models_{\mathsf{SML}^-} \varphi$.*

Soundness It suffices to check that the rules for Δ are sound.

- ($+\Delta$I.). Suppose $\Gamma \models_{\mathsf{SML}^-} +A$. We need to show that $\Gamma \models_{\mathsf{SML}^-} +\Delta A$, i.e. that for all models $\mathcal{M} = \langle \mathcal{W}, \mathcal{V}, R\rangle$ of Γ, for all $w \in \mathcal{W}$, $\mathcal{M}, w \Vdash \Delta A$. So let w and A be arbitrary. We must establish that for all $v \in R(w, A)$, $\mathcal{M}, v \Vdash A$. But since $\mathcal{M}$ is a model of Γ and $\Gamma \models_{\mathsf{SML}^-} +A$ by assumption, we have that for all $v \in \mathcal{W}$, $\mathcal{M}, v \Vdash A$. Thus in particular for all $v \in R(w, A)$, $\mathcal{M}, v \Vdash A$.
- ($+\Delta$E.). Suppose $\Gamma \models_{\mathsf{SML}^-} +\Delta A$. We need to show that $\Gamma \models_{\mathsf{SML}^-} +A$, i.e. that for all models $\mathcal{M} = \langle \mathcal{W}, \mathcal{V}, R\rangle$ of Γ, for all $w \in \mathcal{W}$, $\mathcal{M}, w \Vdash A$. By assumption, $\Gamma \models_{\mathsf{SML}^-} +\Delta A$, so for all models $\mathcal{M}$ of Γ, for all $w \in \mathcal{W}$, $\mathcal{M}, w \Vdash \Delta A$, i.e. for all $v \in R(w, A)$, $\mathcal{M}, v \Vdash A$. Since $w \in R(w, A)$, it follows that $\mathcal{M}, w \Vdash A$. As this holds for every w, it follows that $\Gamma \models_{\mathsf{SML}^-} +A$. □

Completeness This is a standard canonical model construction with a small twist. Let $\Gamma \vdash^s \varphi$ just in case there is a SML^- derivation of φ from Γ which only uses asserted premisses and does not apply ($+\Delta$I.). Call a set of sentences Γ *s-consistent* if $\Gamma \nvdash^s \bot$. Note that there are inconsistent s-consistent sets like $\{+p, +\neg\Delta p\}$ or $\{\ominus p, +p\}$.

Given a consistent set of formulae Γ, let Γ^* be the deductive closure of Γ and let $\mathcal{E} = \{\delta \mid \delta$ is a maximal s-consistent extension of $\Gamma^*\}$, where a superset δ of Γ^* is maximally s-consistent just in case it is s-consistent and all proper supersets of δ are not s-consistent. Since $\vdash^s$ validates *reductio*, for every $\delta \in \mathcal{E}$ and every formula A, either $+A \in \delta$ or $+\neg A \in \delta$ (and never both). Now define the canonical model $\mathcal{M}^\Gamma = \langle \mathcal{W}, \mathcal{V}, R\rangle$ as follows.

- $\mathcal{W} = \mathcal{E}$.
- $v \in R(w, A)$ iff (if $+\Delta A \in w$, then $+A \in v$).
- $\mathcal{V}(w) = \{p \mid +p \in w\}$.

We first show that for all $w \in W$ and sentences A, $w \in R(w, A)$. Suppose there is a counterexample, i.e. w and A such that $w \notin R(w, A)$. By definition, this means that $+\Delta A \in w$ and $+A \notin w$. Since w is a maximal s-consistent extension of Γ^*, this means that $+\neg A \in w$. But then w is not s-consistent, because $+\Delta A \vdash +A$ by ($+\Delta$E.), which, together with $+\neg A$, entails $\bot$. Thus there is no counterexample. Hence $\mathcal{M}^\Gamma$ is an SML^- model.

It is left to show that $\mathcal{M}^\Gamma$ is a model of Γ. We prove by induction on the complexity of sentences A that: $+A \in w$ iff $\mathcal{M}^\Gamma, w \Vdash A$. The cases for atomic A and $A = B \wedge C$ are straightforward, so we only cover negation and the Δ operator.

- If $+\neg A \in w$ then $+A \notin w$, as w is s-consistent. By the induction hypothesis, $\mathcal{M}, w \nVdash A$. Thus $\mathcal{M}, w \Vdash \neg A$.
  Conversely, if $\mathcal{M}, w \Vdash \neg A$, then $\mathcal{M}, w \nVdash A$, so $+A \notin w$ by the induction hypothesis. Since w is a maximally s-consistent set, this means that $+\neg A \in w$.
- Suppose $+\Delta A \in w$. By definition of R, this means that $v \in R(w, A)$ iff $+A \in v$. So, by the induction hypothesis, for all $v \in R(w, A)$, $\mathcal{M}, v \Vdash A$. Thus, $\mathcal{M}, w \Vdash \Delta A$.

Conversely, suppose $\mathcal{M}, w \Vdash \Delta A$. This means that for all $v \in R(v, A)$, $\mathcal{M}, v \Vdash A$. By the induction hypothesis, for all these v, $A \in v$. Now assume that $+\Delta A \notin w$. By the definition of $R(w, A)$ this means that $R(w, A) = \mathcal{W}$. Thus for all $v \in \mathcal{W}$, $A \in v$. But this means that A is a member of all maximally consistent extensions of Γ^*. So it is already a member of Γ^*. Since Γ^* is deductively closed (under the full calculus, including $+\Delta$I.), this means that $+\Delta A \in \Gamma^*$. Since w is a superset of Γ^*, $+\Delta A \in w$. This contradicts the assumption that $+\Delta A \notin w$. By *reductio*, $+\Delta A \in w$.

This concludes the induction. We now show that for all $\varphi \in \Gamma$, $\mathcal{M} \Vvdash \varphi$. If $\varphi = +A$ for some A, then $+A \in \Gamma \subseteq \Gamma^* \subseteq w$ for any $w \in \mathcal{W}$. By the above, $\mathcal{M}, w \Vdash A$. As w is arbitrary, $\mathcal{M} \Vvdash +A$. If $\varphi = \oplus A$, then, since Γ is consistent, $+\neg A \notin \Gamma^*$. Thus there is at least one maximally consistent extension w of Γ^* with $+A \in w$. This means that $\mathcal{M}, w \Vdash A$, so $\mathcal{M} \Vvdash \oplus A$. The case in which $\varphi = \ominus A$ is analogous to the case in which $\varphi = \oplus A$. □

It is easy to construct countermodels to show that, say, $+\Delta\Delta A \vee \Delta\neg\Delta A$ is not a theorem of SML^-. One might worry that **NT** is simply too weak a modal logic. Nevertheless, we contend that having Δ be an **NT** modality *within the multilateral framework* does everything that the supervaluationist needs. This is because the **NT** modal Δ can be embedded under the sign $+$ for *global* truth. The proof of Theorem 2 shows that the following holds (where $\models_g^{\mathbf{NT}}$ denotes the **NT** global consequence relation):

Proposition 3 *For any $\mathcal{L}_\Delta^S$-sentence A and set of $\mathcal{L}_\Delta^S$-sentences Γ,*

$\{+B \mid B \in \Gamma\} \vdash_{\mathsf{SML}^-} +A$ *iff* $\Gamma \models_g^{\mathbf{NT}} A$.

The SML^- logic of assertion is the logic that preserves **NT** truth *in a model*, which is significantly stronger than the preservation of **NT** truth *at a point*.

An example will help. It seems that if one asserts that something is *definitely red and tall*, one ought to be willing to assert that it is *definitely red and tall*.[13] Now the inference from $\Delta A \wedge \Delta B$ to $\Delta(A \wedge B)$ is not valid in **NT**. Nonetheless, if $\Delta A \wedge \Delta B$ is true in an **NT** model (i.e. true at all points), then so is $\Delta(A \wedge B)$. Thus, we have that $+\Delta A \wedge \Delta B \vdash_{\mathsf{SML}^-} +\Delta(A \wedge B)$, as can also be shown as follows.

$$\dfrac{\dfrac{\dfrac{\dfrac{+\Delta A \wedge \Delta B}{+\Delta A}\,(+\wedge\text{E.})}{+A}\,(+\Delta\text{E.}) \qquad \dfrac{\dfrac{+\Delta A \wedge \Delta B}{+\Delta B}\,(+\wedge\text{E.})}{+B}\,(+\Delta\text{E.})}{+A \wedge B}\,(+\wedge\text{I.})}{+\Delta(A \wedge B)}\,(+\Delta\text{I.})$$

By embedding Δ under $+$ one validates inferences than are not available in **NT**. Note that the use of $(+\Delta$I.) in the derivation prevents an application of conditional

[13] This holds not only for assertion, but also for other speech acts such as supposition. In [14], we extend the multilateral framework to cover inferences involving supposition, and the strategy generalises to further speech acts.

proof. Indeed, the *material conditional* $+(\Delta A \wedge \Delta B) \rightarrow \Delta(A \wedge B)$ is not a theorem of SML^{-}. But to explain the validity of inferences such as *if x is definitely red and definitely tall, then x is definitely red and tall*, it is not necessary to derive the corresponding material conditional. The material conditional represents local consequence but, according to the supervaluationist, consequence is global.

Since contraposition is not globally valid, it also does not follow from $+\Delta A \wedge \Delta B \vdash_{\mathsf{SML}^{-}} +\Delta(A \wedge B)$ that $+\neg\Delta A \vee \neg\Delta B$ is derivable from $+\neg\Delta(A \wedge B)$. Indeed, this is not a valid inference in SML^{-}. But there is good reason to think that such inferences are not generally valid. Consider the colour *pink*, whose spectrum is that of a red hue with low saturation. Suppose someone judges that a certain colour is not definitely pink on the grounds that it is not definitely low-saturated and red. Should this entail that they are committed to the colour being not definitely low-saturated or not definitely red? Not necessarily, since it might not be a lack of definite low saturation or a lack of definite redness that underwrote their initial judgement that the color is not definitely pink, but some combination of these factors. Nevertheless, $+\neg\Delta(A \wedge B) \wedge \Delta A \wedge \Delta B$ is provably contradictory in SML^{-}, so one cannot assert that *This colour is not definitely red and low-saturated, but it is definitely red and definitely low-saturated*. This is because when they judge that it is definitely red and definitely low-saturated, there cannot be a lack of definite redness, definite low saturation *or* their combination.

To be sure, one could reconcile the failure of Axioms **4** and **B** with Δ being a normal modal within our system. The easiest way would be to lay down the rule (K) and allow its use under (Weak Inference).

$$\text{(K)}\ \frac{+\Delta(A \rightarrow B) \qquad +\Delta A}{+\Delta B}$$

The rule (K) is already derivable in SML^{-}, but it cannot be used under (Weak Inference), since its derivation involves an application of (+ΔI.). It is straightforward to verify that if (K) can be used under (Weak Inference), Δ is locally a **KT** modality.[14]

However, adding (K) as a primitive inference rule is problematic from the perspective of proof-theoretic semantics. The (K) rule is not obviously an introduction or elimination rule for Δ, but clearly contributes to its meaning. But even setting this issue aside, we submit that it would be inadvisable for the supervaluationist to validate Axiom **K** (and Axiom **E**).

For we have so far merely tackled the basic problem of higher-order vagueness, which was to make it consistent with supervaluationism that *definitely A* is vague. As is well known, this problem can be addressed by giving up Axioms **4** and **B**. However, the gap principle arguments cannot be similarly dealt with. But the upshot of our discussion goes beyond the simple failure of Axioms **4** or **B**. We have traced the problem of higher-order vagueness to one particular rule in SML, in which Δ is an **S5** modality. Dropping this rule invalidates Axioms **4** and **B**, but also results in a non-normal, non-classical definition of Δ. In the next two sections, we show how the non-normality of Δ plays a crucial role in preventing the gap principle arguments from

[14]Cobreros [3] develops the option of treating Δ locally as a **KT** modality under a regional notion of consequence. See also [27, ch. 5].

trivialising the vagueness of Δ and in reconciling supervaluationism with unrestricted higher-order vagueness.[15]

6 How to Reject the Gap Principles

In this and the next section, we show how a straightforward extension of SML^- to a first-order setting is compatible with higher-order vagueness while avoiding the paradoxes generated by gap principles. Our strategy will be to do for higher-order vagueness exactly what the supervaluationist does for ordinary (first-order) vagueness. Thus, the treatment of vague predicates is the same 'all the way up' iterations of Δ.

We start from ordinary vagueness. Consider again the sorites series starting from a_1, who is bald and ending with a_m, who isn't bald. If we let F stand for the predicate *bald*, the following Tolerance Principle is intended to capture the intuition that the transition from baldness to non-baldness is not sharp, that *bald* lacks sharp boundaries.[16]

$$+\forall n(Fa_n \rightarrow Fa_{n+1}) \qquad \text{(Naïve Tolerance)}$$

However, the inductive version of the sorites paradox goes, the Naïve Tolerance Principle and Fa_1 jointly entail Fa_m, thus contradicting the fact that a_m isn't bald.

Some have suggested holding on to the Naïve Tolerance Principle by revising the underlying logic so that the paradoxical conclusion no longer follows or can be tolerated (e.g. [5, 31]). The supervaluationist, for her part, simply takes the Naïve Tolerance Principle to be *false*. All precisifications of a vague predicate F are sharp, i.e. in every precisification there is a threshold k such that Fa_k but $\neg Fa_{k+1}$. This means that the Threshold Principle, the negation of the Naïve Tolerance Principle, is supertrue.

$$+\exists n(Fa_n \wedge \neg Fa_{n+1}) \qquad \text{(Threshold)}$$

But from this it does not follow that there is some particular k such that it is supertrue that k is the cut-off point, since the precisifications disagree on what the witness for the Threshold Principle is. Thus, the supervaluationist maintains that it is *correct to assert* that at some point in a sorites series baldness changes to non-baldness. But it is indeterminate what that point is and so it is *not* correct to assert of any particular point k that it is the cut-off. This, the supervaluationist contends, suffices to explain the intuition that *bald* lacks sharp boundaries.

With respect to the intuitive appeal of the Naïve Tolerance Principle, the *multilateral* supervaluationist can do one better. Stewart Shapiro [21, p. 8] argues that the intuition behind the Naïve Tolerance Principle is best captured by the following principle, which he calls the *principle of tolerance*: if for a vague predicate P, 'two

[15] Similarly, Δ being non-classical is required to account for higher-order vagueness when identity is not vague. See fn. 18 below.

[16] Crispin Wright [29, p. 156] first called attention to tolerance. He called a predicate F *tolerant* with respect to a concept Φ 'if there is also some positive degree of change in respect of Φ insufficient ever to affect the justice with which F applies to a particular case'.

objects a, a_0 … differ only marginally in the relevant respect … then, if one competently judges a to have P, then she cannot judge a_0 not to have P'. We can express this in SML$^-$ as the following axiom schema.

$$\ominus(Fa_k \wedge \neg Fa_{k+1}) \qquad \text{(Bilateral Tolerance)}$$

This means that for any n it is absurd to assert $Fa_n \wedge \neg Fa_{n+1}$. For suppose that for some n, $+(Fa_n \wedge \neg Fa_{n+1})$. Then $\ominus(Fa_n \wedge \neg Fa_{n+1})$ is an instance of the Bilateral Tolerance Principle, from which then $\perp$ follows.

But the Bilateral Tolerance Principle does not entail its naïve counterpart, since the *reductio* meta-inference fails in SML$^-$. One may attempt to reason as follows: assume $+Fa_k$ and for *reductio* that $+\neg Fa_{k+1}$, then by conjunction introduction, $+Fa_k \wedge \neg Fa_{k+1}$ which, as before, contradicts the Bilateral Tolerance Principle. But since the relevant instance of the Bilateral Tolerance Principle is a rejected premiss, the restrictions on ($+\neg$I.) prevent one from discharging $+\neg Fa_{k+1}$ to conclude $+Fa_{k+1}$. We may only use Smileian *reductio* for the discharge to conclude that $\ominus\neg Fa_{k+1}$, and so that $\oplus Fa_{k+1}$. But from $+Fa_k \vdash \oplus Fa_{k+1}$ no paradox can be generated: we cannot continue in the same way from $\oplus Fa_{k+1}$ to derive anything about Fa_{k+2}. The failure of the meta-rules is what allows supervaluationists to accept the Bilateral Tolerance Principle without having to concede its naïve counterpart.

Moving to the first-order setting, existential instantiation joins the ranks of the classical meta-inferences that are not supervaluationistically valid, as observed by Timothy Williamson [28]. The failure of existential instantiation is useful too, since the Threshold Principle and the Bilateral Tolerance Principle are only compatible if existential instantiation fails.[17]

How is the supervaluationist strategy for dealing with the sorites paradox to be extended to the gap principle arguments? The supervaluationist must account for the intuition that the transition from definite baldness to definite not baldness in a sorites sequence is not sharp. Graff Fara argued that, in order to do so, the supervaluationist must accept the F-Gap Principle, which we here formulate as an axiom (instead of a schema) using the resources of first-order logic.

$$+\forall n(\Delta Fa_n \to \neg\Delta\neg Fa_{n+1}) \qquad (F\text{-Gap})$$

Since vagueness cuts at every order, continued Graff Fara, the supervaluationist should also accept the gap principles for any finite iteration of Δ. But the gap principles are incompatible with Δ-Strengthening, thus giving rise to a sorites-like paradox.

[17] Aside from our statement of the Bilateral Tolerance Principle, the failure of existential instantiation also allows the supervaluationist to *assert* that the cut-off points in a sorites series are borderline cases, that is $+\forall n(Fa_n \wedge \neg Fa_{n+1} \to \neg\Delta Fa_n \wedge \neg\Delta\neg Fa_n)$. This claim would be inconsistent with the Threshold Principle if existential instantiation were valid. *Proof sketch*: Assume the Threshold Principle and for existential instantiation that for some k, $+Fa_k \wedge \neg Fa_{k+1}$. If the cut-off points in a sorites series are borderline cases, it follows that $+\neg\Delta Fa_k$. But by ($+\Delta$I.), $+Fa_k$ entails $+\Delta Fa_k$. Contradiction. Discharging the assumption shows a contradiction from the Threshold Principle. Note that the restricted version of existential instantiation we derive below does not validate this argument.

The supervaluationist deals with the sorites paradox by rejecting the Naïve Tolerance Principle. Similarly, she should deal with the higher-order sorites paradox by rejecting the gap principles. Since the strategy takes the same shape at every order of vagueness, we confine our discussion to the F-Gap Principle.

Care is needed, however. For by a classical transformation, the F-Gap Principle is equivalent to the No Δ-Threshold Principle.

$$+\neg\exists n(\Delta Fa_n \wedge \Delta\neg Fa_{n+1}) \quad \text{(No } \Delta\text{-Threshold)}$$

So if the supervaluationist rejects the F-Gap Principle *as false*, she must accept the Δ-Threshold Principle as true.

$$+\exists n(\Delta Fa_n \wedge \Delta\neg Fa_{n+1}) \quad (\Delta\text{-Threshold})$$

But the Δ-Threshold Principle is incompatible with the Borderline Principle, which says that there are borderline cases of baldness.

$$+\exists n(\neg\Delta Fa_n \wedge \neg\Delta\neg Fa_n) \quad \text{(Borderline)}$$

To see that the Δ-Threshold Principle is incompatible with the existence of borderline cases of baldness (and hence with the Borderline Principle), note that such cases must occur after any a_i who is definitely bald and before any a_j who is definitely not bald. But the Δ-Threshold Principle says that there are elements in the sorites series such that one is bald and the one immediately after is not bald. Hence, the supervaluationist cannot accept the Δ-Threshold principle, so cannot reject as false the sentence asserted in the F-Gap Principle.

However, the supervaluationist can reject the sentence asserted in the F-Gap Principle without rejecting it *as false*. In addition, if she adopts the multilateral framework, she can also phrase her rejection of the sentence asserted in the F-Gap Principle as a premiss by using the force-indicator $\ominus$. Thus, she can account for the intuition behind the F-Gap Principle—that the transition from definite baldness to definite not baldness in a sorites sequence is not sharp—by adopting the schematic Δ-Tolerance Principle, which forbids assertions of sentences such as $\Delta Fa_n \wedge \Delta\neg Fa_{n+1}$.

$$\ominus(\Delta Fa_k \wedge \Delta\neg Fa_{k+1}) \quad (\Delta\text{-Tolerance})$$

The situation is analogous to the case of the standard sorites paradox in that both the Naïve Tolerance principle and the gap principles permit marches through a sorities sequence, resulting in paradoxes. In both cases, the supervaluationist blocks the paradox by rejecting the naïve principles and instead adopting versions of these principles that block the march.

The supervaluationist is not yet out of the woods. For there is a difficulty with rejecting the gap principles which did not occur when rejecting the Naïve Tolerance Principle: the Borderline Principle, i.e. the existence of borderline cases of baldness, appears to *entail* the F-Gap Principle.

7 Can the Gap Principles Be Proved?

We have already seen that the Borderline Principle is incompatible with the Δ-Threshold Principle. Zardini [32] has argued that this incompatibility can be established via reasoning that licenses an application of *reductio*. Suppose the Borderline Principle holds. By *reductio*, one may conclude that the negation of the Δ-Threshold Principle holds. But the negation of the Δ-Threshold Principle is the No Δ-Threshold Principle, which is equivalent to the F-Gap Principle. Again, the situation is the same at all orders, so if the F-Gap Principle can be derived in this manner, so can all gap principles.

We will demonstrate that in the natural extension of SML^- to a first-order setting, Zardini's argument fails. At most, the argument can be adapted to show that the Δ-Tolerance Principle follows from accepting the Borderline Principle. To formalise the argument, we first extend SML^- to a first-order language and add rules for the quantifiers. The meaning of the universal quantifier is given by the following rules, which are just the standard rules except that each sentence is prefixed by the assertion sign.

$$(+\forall\text{I.})\ \frac{+A[a/x]}{+\forall x\ A}\ \text{if } a \text{ is any constant symbol not occurring in premisses or undischarged assumptions used to derive } A[a/x]$$

$$(+\forall\text{E.})\ \frac{+\forall x\ A}{+A[t/x]}$$

If we define $\exists x\ A$ as $\neg\forall x\,\neg A$ we can derive (by methods now familiar) the following rules for the existential quantifier.

$$(+\exists\text{E.})\ \frac{+\exists x\ A \qquad \begin{array}{c}[+A[a/x]]\\ +\vdots\\ +B\end{array}}{+B}\ \text{if } a \text{ is any constant symbol not occurring in } A, B, \text{ or premisses or undischarged assumptions, and if } (+\Delta\text{I.}) \text{ was not used to derive } +B.$$

$$(+\exists\text{I.})\ \frac{+A[t/x]}{+\exists x\ A}$$

Among other things, the restrictions on existential instantiation allow the supervaluationist to coherently assert that there is someone who is bald but not definitely bald. Although the $(+\Delta\text{I.})$ rule licenses the inference from baldness to definite baldness, it cannot be applied under the existential quantifier.

We can now proceed to examine Zardini's argument. Consider again a sorites series for *bald*, with the proviso that there is at most one person with the same number of hairs (or, alternatively, that we identify people with exactly the same number of hairs). The series has certain structural features. First, the elements of the series are linearly and ascendingly ordered.

$$+\ \forall m\forall n(a_m = a_n \vee a_m < a_n \vee a_n < a_m) \qquad \text{(Trichotomy)}$$

$$+\ \forall n(a_n < a_{n+1} \wedge \forall m(a_n < a_m \rightarrow (a_m = a_{n+1} \vee a_{n+1} < a_m))) \qquad \text{(Successor)}$$

Second, we have that if someone is definitely bald, then anyone with fewer hairs is also definitely bald, and that if someone is definitely *not* bald, then someone with more hairs is also definitely not bald. We informally used these facts earlier to demonstrate that the Borderline Principle is incompatible with the Δ-Threshold Principle.

We now formalise them as follows.

$$+\ \forall m \forall n (m < n \to (\Delta F a_n \to \Delta F a_m)) \qquad \text{(Monotonicity 1)}$$
$$+\ \forall m \forall n (m < n \to (\Delta \neg F a_m \to \Delta \neg F a_n)) \qquad \text{(Monotonicity 2)}$$

With these principles about the structure of the sorites series on board, it seems that we can now derive the No Δ-Threshold Principle (and hence the F-Gap Principle) from the Borderline Principle. The following is a formal reconstruction of the argument in SML$^-$ extended with quantifier rules.

1.	$+\exists n(\neg\Delta F a_n \wedge \neg\Delta\neg F a_n)$	(Borderline)
2.	$+\exists n(\Delta F a_n \wedge \Delta\neg F a_{n+1})$	(Assumption for $+\neg$I.)
3.	$\quad +\neg\Delta F a_k \wedge \neg\Delta\neg F a_k$	(Assumption for $+\exists$E., 1)
4.	$\quad\quad +\Delta F a_j \wedge \Delta\neg F a_{j+1}$	(Assumption for $+\exists$E., 2)
5.	$\quad\quad\quad +a_k = a_j \vee a_j < a_k \vee a_k < a_j$	($+\forall$E., Trichotomy)
6.	$\quad\quad\quad +a_k = a_j$	(Assumption for $+\vee$E., 5)
7.	$\quad\quad\quad\quad +\Delta F a_j$	($+\wedge$E., 4)
8.	$\quad\quad\quad\quad +\Delta F a_k$	(Intersubstitutivity, 6,7)
9.	$\quad\quad\quad\quad +\neg\Delta F a_k$	($+\wedge$E., 3)
10.	$\quad\quad\quad\quad \bot$	(8, 9)
11.	$\quad\quad\quad +a_j < a_k$	(Assumption for $+\vee$E., 5)
12.	$\quad\quad\quad\quad +a_k = a_{j+1} \vee a_{j+1} < a_k$	(Successor, 11)
13.	$\quad\quad\quad\quad +a_k = a_{j+1}$	(Assumption for $+\vee$E., 12)
14.	$\quad\quad\quad\quad\quad +\Delta\neg F a_{j+1}$	($+\wedge$E., 4)
15.	$\quad\quad\quad\quad\quad +\Delta\neg F a_k$	(Intersubstitutivity, 13, 14)
16.	$\quad\quad\quad\quad\quad +\neg\Delta\neg F a_k$	($+\wedge$E., 3)
17.	$\quad\quad\quad\quad\quad \bot$	(15, 16)
18.	$\quad\quad\quad\quad +a_{j+1} < a_k$	(Assumption for $+\vee$E., 12)
19.	$\quad\quad\quad\quad\quad +\Delta\neg F a_{j+1}$	($+\wedge$E., 4)
20.	$\quad\quad\quad\quad\quad +\Delta\neg F a_k$	(Monotonicity 2, 18, 19)
21.	$\quad\quad\quad\quad\quad +\neg\Delta\neg F a_k$	($+\wedge$E., 3)
22.	$\quad\quad\quad\quad\quad \bot$	(20, 21)
23.	$\quad\quad\quad\quad \bot$	($+\vee$E., 12, 17, 22)
24.	$\quad\quad\quad +a_k < a_j$	(Assumption for $+\vee$E., 5)
25.	$\quad\quad\quad\quad +\Delta F a_j$	($+\wedge$E., 4)
26.	$\quad\quad\quad\quad +\Delta F a_k$	(Monotonicity 1, 24, 25)
27.	$\quad\quad\quad\quad +\neg\Delta F a_k$	($+\wedge$E., 3)
28.	$\quad\quad\quad\quad \bot$	(26, 27)
29.	$\quad\quad\quad \bot$	($+\vee$E., 5, 10, 23, 28)
30.	$\quad\quad \bot$	($+\exists$E., 2, 4, 29)
31.	$\quad \bot$	($+\exists$E, 1, 3, 30)
32.	$+\neg\exists n(\Delta F a_n \wedge \Delta\neg F a_{n+1})$	($+\neg$I., 2, 31)

Although the argument uses meta-inferences like *reductio* and existential instantiation that are not generally valid, they are properly applied here, since the argument contains only asserted premisses and involves no application of the ($+\Delta$I.) rule. Thus, there seems to be little room for maneuvering. However, as Zardini [32, p. 31] himself points out, one could reject the intersubstitutivity of identicals within Δ-contexts. In that case, its application at lines (8) and (15) in the derivation would be faulty. We now show that the failure of the intersubstitutivity of identicals within Δ-contexts naturally arises when one extends the supervaluationist multilateral framework with suitable identity rules. So the argument fails to establish the No Δ-Threshold Principle.

The usual elimination rule for identity is the intersubstitutivity of identicals in transparent contexts (i.e., for present purposes, not in the scope of any Δ). Stephen Read [19] formulates an introduction rule for identity with which the usual elimination role can be shown to be in harmony. We can adopt these rules within the multilateral framework simply by prefixing each of the sentences occurring in them with the assertion sign.

$$(+=\text{I.})\ \frac{\begin{matrix}[+Fa] \\ \vdots \\ +Fb\end{matrix} \qquad \begin{matrix}[+Fb] \\ \vdots \\ +Fa\end{matrix}}{+a=b}$$ where F is a predicate symbol not occurring in premisses or undischarged assumptions

$$(+=\text{E.}_1)\ \frac{+a=b \quad +Fa}{+Fb}$$ where F is a predicate symbol $$(+=\text{E.}_2)\ \frac{+a=b \quad +Fb}{+Fa}$$ where F is a predicate symbol

The identity rules immediately yield the principle of intersubstitutivity of identicals in transparent contexts, i.e. $+\forall x \forall y(x = y \rightarrow (Fx \leftrightarrow Fy))$ where F is any predicate symbol. But this does not entail the principle in *Δ-contexts*.

Although $+\forall x \forall y(x = y \rightarrow (Fx \leftrightarrow Fy))$ delivers $+\Delta\forall x \forall y(x = y \rightarrow (Fx \leftrightarrow Fy))$ by an application of the ($+\Delta$I.) rule, the non-normality of Δ prevents one from concluding that $+\forall x \forall y(\Delta x = y \rightarrow (\Delta Fx \leftrightarrow \Delta Fy))$.[18] Nonetheless, the *rule* version of the intersubstitutitivity of identicals in basic Δ-contexts is derivable by (repeatedly) applying ($+\Delta$E.), ($+=$E.) and ($+\Delta$I.).

$$(\Delta^n\text{-Intersubstitutivity})\ \frac{+a=b \qquad +\Delta\Delta^n Fa}{+\Delta\Delta^n Fb}$$ where F is a predicate symbol

Since its derivation uses the ($+\Delta$I.) rule, however, the Δ-Intersubstitutivity Rule cannot be applied in steps (8) and (15) of Zardini's argument. For the ($+\Delta$I.) rule is disallowed in *reductio* arguments. The intersubstitutivity of identicals in formulae involving complex combinations of Δ and the other connectives is not easily captured in a single derived rule (see [1] for related discussion).

The rule version of the intersubstitutivity of identicals in Δ-contexts suffices to account for the validity of inferences such as the one from *Superman is definitely strong* and *Superman is Clark Kent* to *Clark Kent is definitely strong*. Such

[18] If we accept that identity is rigid across precisifications (so that identity is not vague), Axiom **E** already suffices to establish the principle of intersubstitutivity of identicals in Δ-contexts. So the fact that Δ is non-classical (besides being non-normal) matters here. *Proof:* Suppose $a = b$ is true at a point. By assumption, $a = b$ is true at all points, so Fa and Fb are extensionally equivalent. By **E**, ΔFa and ΔFb are extensionally equivalent. But this just means that $a = b \rightarrow (\Delta Fa \leftrightarrow \Delta Fb)$ is true at every point.

inferences are globally valid and hence, from a supervaluationist standpoint, valid *tout court*, since they preserve truth *simpliciter*. What cannot be had is the principle $+\forall x\forall y((x = y \wedge \Delta Fx) \rightarrow \Delta Fy)$, which states that the relevant inferences are not only globally but also *locally* valid. Thus, pending an argument that good reasoning involving vague terms requires not only the validity *tout court* of intersubstitutivity into Δ-contexts, but its local validity, the Δ-intersubstitutivity principle $+\forall x\forall y((x = y \wedge \Delta Fx) \rightarrow \Delta Fy)$ remains unjustified.

Thus, Zardini's proposed derivation of the No Δ-Threshold Principle from the Borderline Principle is blocked in a natural extension of SML^- to first-order logic with identity.[19] To show that no such derivation exists, we develop a model theory for quantified SML^- (QSML^- for short), the deductive system which results from adding to SML^- the above rules for the universal quantifier and identity.

QSML^- is cast in a language $\mathcal{L}^S_{\Delta,=}$ obtained by extending $\mathcal{L}^S_\Delta$ in the obvious way with quantifiers, identity and the necessary non-logical vocabulary (countably many constant symbols including 0, function $+1$, relation $<$ and a predicate F). The notion of a model is defined as follows.

Definition 5 A QSML^--model is a tuple $\mathcal{M} = \langle \mathcal{W}, \mathcal{D}, \mathcal{I}, R\rangle$, where $\mathcal{W}$ is a set of points, $\mathcal{D}$ is a function mapping each $w \in \mathcal{W}$ to a domain $\mathcal{D}(w)$, $\mathcal{I}$ is a function mapping each w to a function $\mathcal{I}(w)$ that interprets the non-logical vocabulary in the usual way, and $R : \mathcal{W} \times \mathcal{L}^S_{\Delta,=} \rightarrow \mathcal{P}(\mathcal{W})$ is a function mapping every point and $\mathcal{L}^S_{\Delta,=}$-sentence to a set of points such that for all $w \in \mathcal{W}$ and A, $w \in R(w, A)$.

Local satisfaction is defined as follows.

- $\mathcal{M}, w \Vdash P(t_1, ..., t_n)$ iff $(t_1^{\mathcal{I}(w)}, ..., t_n^{\mathcal{I}(w)}) \in \mathcal{I}(w)(P)$, where for a term t, $t^{\mathcal{I}(w)}$ is its evaluation given the interpretation $\mathcal{I}(w)$ (defined in the usual way).
- Boolean connectives as usual.
- $\mathcal{M}, w \Vdash \Delta A$ iff for all $v \in R(w, A)$, $\mathcal{M}, v \Vdash A$.
- $\mathcal{M}, w \Vdash \forall x\, A$ iff for all terms t, $\mathcal{M}, w \Vdash A[t/x]$.
- $\mathcal{M}, w \Vdash t_1 = t_2$ iff $t_1^{\mathcal{I}(w)} = t_2^{\mathcal{I}(w)}$.

The clause for the universal quantifier is the substitutional one, not the usual Tarskian one. In Tarskian model theory, $\forall x\, A$ is defined to be true relative to a

[19] Another strategy to block the argument, similarly germane to our general approach, is to reject the Monotonicity Principles and give due to their intuitive appeal by validating the corresponding rules. We have seen that it is far from obvious what the correct meaning of Δ should be, particularly when embedded in complex formulae. Hence, ideally, principles involving Δ such as the Monotonicity Principles should be justified by deriving them from more basic assumptions. Now it is plausible that it is part of the meaning of the predicate F standing for *baldness* that for all m, n it is the case that $+m < n \rightarrow (Fa_n \rightarrow Fa_m)$ and that $+m < n \rightarrow (\neg Fa_m \rightarrow \neg Fa_n)$. Being Δ-free, the use of the material conditional to formulate these assumptions is not problematic. Using these assumptions, we can derive the rule versions of the Monotonicity Principles.

$$\text{(M1)}\ \frac{+i < j \qquad +\Delta Fj}{+\Delta Fi} \qquad \text{(M2)}\ \frac{+i < j \qquad +\Delta\neg Fi}{+\Delta\neg Fj}$$

However, the derivation involves an application of ($+\Delta$I.) and so the Monotonicity Principles themselves do not follow. As in the case of the intersubstitutivity of identicals, the non-normality of Δ is crucial here.

variable assignment f just in case A is true for all assignments f' that differ from f only in what they assign to x. We do not require assignment functions to define satisfaction (notably, this means that sentences with open variables do not have satisfaction conditions) and instead substitute different terms in A.[20]

Global satisfaction is defined as in the propositional case:

- $\mathcal{M} \Vdash\!\models +A$ iff for all $w \in \mathcal{W}$, $\mathcal{M}, w \Vdash A$.
- $\mathcal{M} \Vdash\!\models \oplus A$ iff for some $w \in \mathcal{W}$, $\mathcal{M}, w \Vdash A$.
- $\mathcal{M} \Vdash\!\models \ominus A$ iff for some $w \in \mathcal{W}$, $\mathcal{M}, w \nVdash A$.

QSML$^-$ is sound and complete with respect to this model theory. Write $\Gamma \models_{\mathsf{QSML}^-} \varphi$ iff for all QSML$^-$-models $\mathcal{M}$ with $\mathcal{M} \Vdash\!\models \psi$ for all $\psi \in \Gamma$, it is the case that $\mathcal{M} \Vdash\!\models \phi$.

Theorem 3 *For any $\mathcal{L}^S_{\Delta,=}$-formula φ and set of $\mathcal{L}^S_{\Delta,=}$-formulae Γ, $\Gamma \vdash_{\mathsf{QSML}^-} \varphi$ iff $\Gamma \models_{\mathsf{QSML}^-} \varphi$.*

Soundness The soundness of the connective and Δ rules carries over from the propositional case and the identity rules are easy to verify. So we only need to check the universal quantifier rules.

- (+∀E.). Suppose $\mathcal{M}, w \Vdash \forall x\, A$ and let t_0 be any term. We need to show that $\mathcal{M}, w \Vdash A[t_0/x]$. By definition, if $\mathcal{M}, w \Vdash \forall x\, A$, then $\mathcal{M}, w \Vdash A[t/x]$ for any t, including t_0, so we are done.[21]
- (+∀I.). Suppose the last step in some derivation D is an application of (+∀I.) to infer $+\forall x\, A$ from $+A[a/x]$ where a does not occur in premisses or undischarged assumptions.

  Let $\{\varphi_i \mid i < n\}$ be the finite set of premisses and undischarged assumptions used in the derivation of $+A[a/x]$. So there is a subderivation D' of D such that:

  $$\{\varphi_i \mid i < n\} \vdash^{D'} +A[a/x].$$

  Let t be any term. By assumption, a does not occur anywhere in the φ_i, so the derivation D^t that is obtained by uniformly substituting t for a in D' is a derivation *from the same premisses* (as they remain unchanged by the substitution).

[20] The substitutional definition of universal quantification allows models where the universal quantifier has a nonstandard interpretation. That is, there are $\mathcal{M}$ and w such that there is an $d \in D(w)$ with $\mathcal{M}, w \Vdash \forall x\, F(x)$ but $d \notin \mathcal{I}(w)(F)$. This can happen when d is not denoted by any term. One may rule out nonstandard interpretations by requiring that models be *standard*, where a model is standard if at every point w, every $d \in \mathcal{D}(w)$ is denoted by some term according to $\mathcal{I}(w)$. The proof of Theorem 3 shows that QSML$^-$ is sound and complete with respect to both the set of standard QSML$^-$-models and the set of all QSML$^-$-models. This is a special case of the result that the natural deduction rules for the quantifiers are not categorical [2].

[21] This is where the substitutional definition of ∀ is important. If $A = \Delta Fx$ then we must show that $\mathcal{M}, w \Vdash \Delta Ft$. If ∀ were defined in the Tarskian way by x-variant assignments, $\mathcal{M}, w, f \Vdash \forall x\, \Delta Fx$ would mean that for all x-variants f' of f, $\mathcal{M}, w, f' \Vdash \Delta Fx$. But if $R(w, Ft) \neq R(w, Fx)$, then this does not imply that $\mathcal{M}, w, f \Vdash \Delta Ft$. Ultimately, this is because ΔA is defined by making use of *syntactic* properties of A, so the syntactic forms of the instances of a quantified sentence matter for its interpretation. Thus, when interpreting ∀, one must track its instances syntactically as well.

Therefore, $\{\varphi_i \mid i < n\} \vdash +A[t/x]$. Note that D' is shorter than D, so D^t is too. Hence by the induction hypothesis, $\{\varphi_i \mid i < n\} \models_{\mathsf{QSML}^-} +A[t/x]$. This holds for any t, so $\{\varphi_i \mid i < n\} \models_{\mathsf{QSML}^-} +\forall x\, A$. $\square$

Completeness follows from a straightforward adaption of the usual term model construction to the method of using maximally s-consistent sets instead of maximally consistent sets (see the completeness proof for SML^-).

On the basis of this model theory, we can easily see that substituting identicals in Δ-contexts is not locally valid. For if $R(w, Fa) \neq R(w, Fb)$, it may be the case that $\mathcal{M}, w \Vdash a = b$ and $\mathcal{M}, w \Vdash \Delta Fa$ while $\mathcal{M}, w \nVdash \Delta Fb$. In the model theory, $a = b$ may be locally but not globally true. This is *also* a reason why it is not locally valid to substitute identicals in Δ-contexts. But this is tangential to our point. Even if one insisted that identity is rigid across precisifications, substituting identicals in Δ-contexts would not be locally valid. For even if it is *globally* the case that $a = b$ (i.e. all w agree that $a^{\mathcal{I}(w)} = b^{\mathcal{I}(w)}$), it may still be the case for some w that $R(w, Fa) \neq R(w, Fb)$ because a and b are *syntactically* distinct regardless of their interpretation.[22]

This feature of the logic lends itself to a contextualist interpretation.[23] Two sentences *a is definitely tall* and *b is definitely tall* may be precisified to different standards of definite tallness (i.e. it may be that $R(w, Ta) \neq R(w, Tb)$). This is plausible, for instance, when a is a building and b is a person, since different standards of (definite) tallness apply. More dramatically, this can also happen when a and b co-refer. We submit that this is still plausible if vagueness is context-sensitive, since the way an object is referred to can affect the relevant standards. One may refer to, say, Jeremy Lin as *that person* or as *that basketball player*. These two ways of referring are associated with different standards of (definite) tallness. Formally, this is captured by the fact that (locally) the accessibility relations associated with *that person is tall* and *that basketball player is tall* can differ.

The local failure of the intersubstitutivity of identicals in Δ-contexts can be exploited to show that the No Δ-Threshold Principle is not derivable from the Borderline Principle, even in the presence of the Trichotomy, Successor and Monotoncity Principles.

$$+\ \exists n(\neg\Delta Fa_n \wedge \neg\Delta\neg Fa_n) \qquad \text{(Borderline)}$$

$$+\ \neg\exists n(\Delta Fa_n \wedge \Delta\neg Fa_{n+1}) \qquad \text{(No } \Delta\text{-Threshold)}$$

It is easy to construct a QSML^- model $\mathcal{M}$ such that Borderline Principle is satisfied in $\mathcal{M}$, but No Δ-Threshold is not—that is, a model such that for all w there is a constant a_k such that $\mathcal{M}, w \Vdash \neg\Delta Fa_k \wedge \neg\Delta\neg Fa_k$ and there are some v and a_j such that $\mathcal{M}, v \Vdash \Delta Fa_j \wedge \Delta\neg Fa_{j+1}$. It remains to show that the Trichotomy, Successor

[22]But if $a = b$ and ΔFa are both globally true, then so is ΔFb. This is due to Axiom **T**. If ΔFa holds at some w, Fa holds at w. So if ΔFa is globally true, then so is Fa and hence Fb. So there are no points at which Fb is not true and so ΔFb cannot be false anywhere. Proof-theoretically, this is the derived rule of Δ-Intersubstitutivity.

[23]Shapiro [21] develops a theory of vagueness which combines contextualist and supervaluationist elements. We hope to explore the similarities between our view and Shapiro's in future work.

and Monotonicity Principles can also hold at all w (in particular at v). If M, v satisfies the the Trichotomy Principle, then

$$M, v \Vdash a_k = a_j \vee a_k < a_j \vee a_j < a_k.$$

By the Successor Principle, this means that

$$M, v \Vdash a_k = a_j \vee a_k < a_j \vee a_{j+1} = a_k \vee a_{j+1} < a_k.$$

And if M, v satisfies the Monotonicity Principles, then $a_k < a_j$ and $a_{j+1} < a_k$ lead to contradictions. It follows that

$$M, v \Vdash a_k = a_j \vee a_k = a_{j+1}.$$

But no contradiction follows from this, since $a_k = a_j \wedge \neg\Delta Fa_k \wedge \Delta Fa_j$ is satisfiable, as we can define a model such that $R(v, Fa_k) \neq R(v, Fa_j)$. For similar reasons, there are models satisfying $a_k = a_{j+1} \wedge \neg\Delta\neg Fa_k \wedge \Delta\neg Fa_{j+1}$ as well.[24]

Hence the supervaluationist can assert that there are borderline cases *and* reject the F-Gap Principle. Which is what she should do. Now the countermodel construction showing that the No Δ-Threshold principle does not follow from the Borderline Principle makes it clear that the following nonetheless holds:

$$+\forall m \forall n(\neg\Delta Fa_m \wedge \neg\Delta\neg Fa_m \wedge \Delta Fa_n \wedge \Delta\neg Fa_{n+1}) \rightarrow a_m = a_n \vee a_m = a_{n+1}.$$

That is, if a_m is a borderline case of baldness and n is a threshold, then a_m is either a_n or the successor of a_n. While the conclusion appears absurd, our supervaluationist is untroubled, since she does not accept the existence of a threshold, that is of an n with $\Delta Fa_n \wedge \Delta\neg Fa_n$. This does not mean that she rejects the existence of a threshold *as false*—this would entail the F-Gap Principle. Instead she *weakly* rejects the existence of a threshold, as the Δ-Tolerance Principle demands.

$$\ominus(\Delta Fa_n \wedge \Delta\neg Fa_{n+1}) \qquad (\Delta\text{-Tolerance})$$

In fact, suppose that in the model theory we require that all points have a shared domain and agree on the interpretations of terms (so that, roughly, identity is not vague). Proof-theoretically, this means that $+a = b$ is derivable from $\oplus a = b$. Then Zardini's argument can be adapted to show that the Borderline, Trichotomy, Successor and Monotonicity Principles jointly *entail* all instances of the Δ-Tolerance Principle, i.e. that for any n, $\ominus\Delta Fa_n \wedge \Delta\neg Fa_{n+1}$. The proof is in the Appendix.

The Δ-Tolerance Principle suffices to account for the apparent lack of a sharp boundary between *definitely bald* and *definitely not bald* without sanctioning the gap principles. Earlier we told the following story about the Naïve Tolerance Principle on behalf of the supervaluationist. The phenomenon of ordinary vagueness might lead one to assume that if someone with j hairs on their head is bald, then someone with $j + 1$ hairs is bald. But on closer inspection, so the story goes, the phenomenon only sanctions the Tolerance Principle: it is never correct to assert that someone with j hairs is bald and someone with $j + 1$ hairs is not. We can now tell a similar story

[24] Although Δ is, locally, a very weak modality, it is not the case that *anything* goes. For example, a formula like $a = b \wedge \Delta Fa \wedge \Delta\neg Fb$ is not satisfiable, since ΔFa entails Fa and $\Delta\neg Fb$ entails $\neg Fb$, which are incompatible with $a = b$. But we are never in such a case here.

about the F-Gap Principle. The phenomenon of higher-order vagueness might lead one to assume that if someone with j hairs is definitely bald, then someone with $j+1$ hairs is not definitely not bald. But the phenomenon only sanctions the Δ-Tolerance Principle: it is never correct to assert that someone with j hairs is definitely bald and someone with $j+1$ hairs is definitely not bald.

This story can be extended to any order of vagueness: we reject the n^{th}-order version of the Gap Principle and account for the phenomenon of vagueness at every order by accepting the schematic Δ^n-Tolerance Principle.

$$\ominus(\Delta\Delta^n Fa_k \wedge \Delta\neg\Delta^n Fa_{k+1}) \qquad (\Delta^n\text{-Tolerance})$$

It may be objected that the F-Gap Principle is central to our practices involving vague terms in a way that the Naïve Tolerance Principle isn't: even if it is wrong to infer Fa_{j+1} from Fa_j, it should be possible to infer $\neg\Delta\neg Fa_{j+1}$ from ΔFa_j. The Bilateral Tolerance Principle, the objection goes, may be an adequate replacement for its naïve counterpart, but the Δ^n-Gap Principle is needed to account for good reasoning involving vague terms. However, for this objection to succeed, it must be established not only that good reasoning involving vague terms requires the validity of the inferences from $+\Delta\Delta^n Fa_j$ to $+\neg\Delta\neg\Delta^n Fa_{j+1}$, but that it requires the assertion of the corresponding material conditionals. This is because we could extend QSML$^-$ with the *rule* versions of the gap principles, but disallow their use under (Weak Inference).

$$\text{(Gap Rule)}\ \frac{+\Delta\Delta^n Fa_k}{+\neg\Delta\neg\Delta^n Fa_{k+1}}$$

Given that it does not satisfy the standard proof-theoretic constraints on the admissibility of rules of inference—for a start, Δ occurs in both the premiss and the conclusion—the Gap Rule is best understood as a non-logical rule, which does not contribute to the meaning of Δ. Adding the Gap Rule without permitting its use under (Weak Inference) means that it cannot be contraposed, which blocks the gap principle arguments (see Section 4). Since the Gap Rule cannot be used in existential instantiation arguments either, it is also compatible with the Borderline Principle. Disaster only results from accepting the material conditionals $+\Delta\Delta^n Fa_j \rightarrow \neg\Delta\neg\Delta^n Fa_{j+1}$. But given the failure of conditional proof, we may accept the Gap Rule without accepting its material conditional version. Thus, even if good reasoning involving vague terms requires the inferences from $+\Delta\Delta^n Fa_j$ to $+\neg\Delta\neg\Delta^n Fa_{j+1}$ to be valid, this can be accommodated within our account. It is only the corresponding material conditionals which cannot be accepted.

Are arguments for accepting these material conditionals going to be forthcoming? We cannot rule this out in principle, but would like to stress just how strong the conclusion of any such argument must be. In the simplest case, it would have to establish the conditional $+\Delta Fa_j \rightarrow \neg\Delta\neg Fa_{j+1}$, which means that it is absurd for anyone to refrain from believing *the person with j hairs is not definitely bald or the person with one hair more isn't definitely not bald*. The same holds for the relevant material conditionals of higher order. The Gap Rule has no such consequence, since its application requires a previous assertion about definite baldness in the local proof context.

8 Conclusion

Global consequence has the reputation of being proof-theoretically intractable. The failure of the classical meta-inferences has sometimes be taken to imply that no proof system for global consequence is forthcoming—or, at any rate, that any such proof system would be too complex to provide a regimentation of canons of reasoning which can be feasibly used in actual deductive practice. We have shown that such concerns are ill-founded. The multilateral system SML consists of simple rules formulated in terms of assertion, rejection and weak assertion, and is sound and complete with respect to standard supervaluationist model theory under a global notion of validity. In addition, all inference rules of SML meet the standard proof-theoretic constraints on their admissibility.

Our proof system, furthermore, reveals a natural solution to the problem of higher-order vagueness. It is often recognized that the basic problem requires rejecting Axioms **4** or **B** for Δ. However, proof analysis shows that it is one particular rule in SML that makes it impossible for ΔA to be borderline. But forfeiting this rule results in the failure not only of both Axiom **4** *and* Axiom **B**, but *also* of Axioms **K** and **E**: Δ becomes a non-normal, non-classical modality. In the extension of the reduced calculus with rules for quantifiers and identity, Zardini's derivation of the gap principles fails. The failure of this derivation allows supervaluationists to treat vagueness uniformly at every order: they can reject the principles used to traverse a sorites sequence.

The multilateral calculus, therefore, enables supervaluationists to specify which arguments patterns govern good reasoning and to provide a proof analysis leading to the rejection of Axioms **K** and **E**. But the calculus offers further benefits to supervaluationists: by incorporating rejections in its object language, it provides a formal framework to articulate and examine the consequences of the strategy of rejecting a sentence without rejecting it as false.[25] In particular, using rejections, the supervaluationist can formulate tolerance principles for any order. The failure of *reductio* and other classically valid meta-inferences plays an important role in this approach, since in the presence of these inferences, the tolerance principles entail paradoxical gap principles. Thus, far from being troublesome, the failure of the relevant meta-inferences is a central component of the supervaluationist strategy.

We started from a standard understanding of supervaluationism, which is usually implemented using the tools of model theory. We have shown that proof theory has much to offer to supervaluationism. However, there is also nothing in our proof theory that implies that vagueness is a matter of semantic indecision, instead of an epistemic matter. Whether our approach points to a reconciliation between traditional supervaluationism and epistemicism about vagueness is a question that we hope to tackle in future work.

[25] If we understand a rejection of A as expressing that A is subfalse (i.e. there being at least one precisification where A fails), then some versions of supervaluationism (notably, the one in which Δ is an **S5** modality) can express rejection as $\neg\Delta$. However, if Δ is (locally) an **NT** modality—as we have argued it should be—rejection cannot be reduced to $\neg\Delta$.

Appendix

We prove that the instances of the Δ-Tolerance Principle (i.e. $\ominus\Delta Fa_n \wedge \Delta\neg Fa_{n+1}$ for any n) are derivable from the Borderline, Trichotomy, Successor and Monotonicity Principles in QSML^- augmented with the Rigidity Rule. The Rigidity Rule says that identity is rigid across precisifications.

$$\text{(Rigidity)}\ \frac{\oplus a = b}{+a = b}$$

For the proof, we require the following derived rules; note that $(+\Delta\text{I.})$ may be used in the subderivations of $(\oplus\exists\text{E.})$ and $(\oplus\vee\text{E.})$.

$$(\oplus\exists\text{E.})\ \frac{\oplus\exists x A \qquad \begin{matrix}[\oplus A[a/x]]\\ \vdots \\ \bot\end{matrix}}{\bot}\ \text{if } a \text{ is any constant symbol not occurring in } A,\ B, \text{ or premisses or undischarged assumptions.}$$

$$(\oplus\vee\text{E.})\ \frac{\oplus A\vee B \qquad \begin{matrix}[\oplus A]\\ \vdots \\ \bot\end{matrix} \qquad \begin{matrix}[\oplus B]\\ \vdots \\ \bot\end{matrix}}{\bot} \qquad (\oplus\wedge\text{I.})\ \frac{+A \qquad \oplus B}{\oplus A\wedge B} \qquad (\oplus\wedge\text{E.})\ \frac{\oplus A\wedge B}{\oplus A}$$

Proof of $(\oplus\exists\text{E.})$:

$$\dfrac{\dfrac{\dfrac{\oplus\exists x A}{\oplus\neg\forall x\neg A}\,\text{(Abbr.)}}{\ominus\forall x\neg A}\,(\oplus\neg\text{E.}) \qquad \dfrac{\dfrac{\dfrac{\dfrac{[\ominus\neg A[a/x]]^1}{\oplus A[a/x]}\,(\ominus\neg\text{E.})}{\bot}\,\text{(by assumption)}}{+\neg A[a/x]}\,(\text{SR}_2)^1}{+\forall x\neg A}\,(+\forall\text{I.}),\ a \text{ does not occur in premisses or undischarged assumptions}}{\bot}\,\text{(Rejection)}$$

Proof of $(\oplus\vee\text{E.})$:

$$\dfrac{\dfrac{\dfrac{\oplus A\vee B}{\oplus\neg(\neg A\wedge\neg B)}\,\text{(Abbr.)}}{\ominus\neg A\wedge\neg B}\,(\oplus\neg\text{E.}) \qquad \dfrac{\dfrac{\dfrac{\dfrac{[\ominus\neg A]^1}{\oplus A}\,(\ominus\neg\text{E.})}{\bot}\,\text{(by assumption)}}{+\neg A}\,(\text{SR}_2)^1 \qquad \dfrac{\dfrac{\dfrac{[\ominus\neg B]^2}{\oplus B}\,(\ominus\neg\text{E.})}{\bot}\,\text{(by assumption)}}{+\neg B}\,(\text{SR}_2)^2}{+\neg A\wedge\neg B}}{\bot}\,\text{(Rejection)}$$

The derivations of the $(\oplus\wedge)$-rules are straightforward applications of (Weak Inference) and the $(+\wedge)$-rules.

Now recall that from Zardini's proof we get the following.

(Z) $\quad +\forall m\forall n(\neg\Delta Fa_m \wedge \neg\Delta\neg Fa_m \wedge \Delta Fa_n \wedge \Delta\neg Fa_{n+1}) \rightarrow a_m = a_n \vee a_m = a_{n+1}.$

Then note that $(*)$ $\oplus\ (a_j = a_i), +\Delta Fa_i \wedge \Delta\neg Fa_{i+1}, \oplus\neg\Delta Fa_j \wedge \neg\Delta\neg F_a j \vdash \bot.$

$$\dfrac{\dfrac{\dfrac{[\oplus a_j = a_i]^3}{+a_j = a_i}\,\text{(Rigidity)} \qquad \dfrac{+\Delta Fa_i\wedge\Delta\neg Fa_{i+1}}{+\Delta Fa_i}\,(+\wedge\text{E.})}{+\Delta F_j}\,(\Delta\text{-Intersubstitutivity}) \qquad \dfrac{\dfrac{\oplus\neg\Delta Fa_j\wedge\neg\Delta\neg Fa_i}{\oplus\neg\Delta Fa_j}\,(\oplus\wedge\text{E.})}{\ominus\Delta Fa_j}\,(\ominus\neg\text{E.})}{\bot}\,\text{(Rejection)}$$

By an analogous proof, $(*')$ $\oplus(a_j = a_{i+1}), +\Delta Fa_i \wedge \Delta\neg Fa_{i+1}, \oplus\neg\Delta Fa_j \wedge \neg\Delta\neg Fa_j \vdash \bot$. Then derive any instance of the Δ-Tolerance Principle as follows.

$$\dfrac{\dfrac{\dfrac{\text{(Borderline)}}{+\exists n(\neg\Delta Fa_n\wedge\neg\Delta\neg Fa_n)}}{\oplus\exists n(\neg\Delta Fa_n\wedge\neg\Delta\neg Fa_n)}\text{(Assertion)} \quad \dfrac{\dfrac{\dfrac{[+\Delta Fa_i\wedge\Delta\neg Fa_{i+1}]^1 \quad [\oplus\neg\Delta Fa_j\wedge\neg\Delta\neg Fa_j]^2}{\oplus\Delta Fa_i\wedge\Delta\neg Fa_{i+1}\wedge\neg\Delta Fa_j\wedge\neg\Delta\neg Fa_j}(\oplus\wedge\text{I.})}{\oplus a_j=a_i\vee a_j=a_{i+1}}(\text{Z}) \quad \dfrac{[\oplus a_j=a_i]^3}{\bot}(*) \quad \dfrac{[\oplus a_j=a_{i+1}]^3}{\bot}(*')}{\bot}(\oplus\vee\text{E.})^3}{\dfrac{\bot}{\ominus\Delta Fa_i\wedge\Delta\neg Fa_{i+1}}(\text{SR}_1)^1}(\oplus\exists\text{ E.})^2$$

The argument requires the use of ($\oplus\exists$E.) and ($\oplus\vee$E.) since the proofs of $\bot$ from $\oplus(a_j = a_i)$ and $\oplus(a_j = a_{i+1})$ require the Δ-Intersubstitutivity Rule, which is not licit under their counterparts with $+$. For the same reason, one cannot apply *reductio* in the final step of the proof to derive the F-Gap Principle.

Acknowledgements This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 758540) within the project *From the Expression of Disagreement to New Foundations for Expressivist Semantics*. We would like to thank two anonymous reviewers for very helpful comments. An earlier version of this paper was presented at the Logic Colloquium of the Munich Center for Mathematical Philosophy. We are grateful to the members of the audience for their valuable feedback.

References

1. Aloni, M. (2005). Individual concepts in modal predicate logic. *Journal of Philosophical Logic*, *34*, 1–64.
2. Bonnay, D., & Westerståhl, D. (2016). Compositionality solves Carnap's problem. *Erkenntnis*, *81*, 721–739.
3. Cobreros, P. (2008). Supervaluationism and logical consequence: a third way. *Studia Logica*, *90*, 291–312.
4. Cobreros, P. (2011). Varzi on supervaluationism and logical consequence. *Mind*, *120*, 833–843.
5. Cobreros, P., Egré, P., Ripley, D., & van Rooij, R. (2012). Tolerant, classical, strict. *Journal of Philosophical Logic*, *41*, 347–385.
6. Cobreros, P., & Tranchini, L. (2019). Supervaluationism, subvaluationism and the sorites paradox. In S. Oms, & E. Zardini (Eds.) *The sorites paradox,* pp. 38–62. Cambridge University Press, Cambridge.
7. Dicher, B. (2020). Hopeful monsters: a note on multiple conclusions. *Erkenntnis*, *85*, 77–98.
8. Dummett, M. (1991). *The logical basis of metaphysics.* Harvard University Press, Cambridge MA.
9. Fine, K. (1975). Vagueness, truth, and logic. *Synthese*, *30*, 265–300.
10. Graff Fara, D. (2003). Gap principles, penumbral consequence, and infinitely higher-order vagueness. In J. C. Beall (Ed.) *Liars and heaps: new essays on paradox,* pp. 195–221. Clarendon Press, Oxford.

11. Graff Fara, D. (2010). Scope confusions and unsatisfiable disjuncts: Two problems for supervaluationism. In R. Dietz, & S. Moruzzi (Eds.) *Cuts and clouds: vagueness, its nature, and its logic,* pp. 373–382. Oxford University Press, Oxford.
12. Incurvati, L., & Schlöder, J. J. (2017). Weak rejection. *Australasian Journal of Philosophy*, *95*, 741–760.
13. Incurvati, L., & Schlöder, J. J. (2019). Weak assertion. *Philosophical Quarterly*, *69*, 741–770.
14. Incurvati, L., & Schlöder, J. J. (forthcoming). Epistemic multilateral logic. *Review of Symbolic Logic*.
15. Jones, N. K. (2011). Williams on supervaluationism and logical revisionism. *Journal of Philosophy*, *108*, 633–641.
16. Keefe, R. (2000). *Theories of vagueness*. Cambridge University Press, Cambridge.
17. Lewis, D. (1973). *Countefactuals.* Harvard University Press, Cambridge MA.
18. Lewis, D. (1986). *On the plurality of worlds.* Blackwell, Oxford.
19. Read, S. (2004). Identity and harmony. *Analysis*, *64*, 113–119.
20. Rumfitt, I. (2000). 'Yes' and 'No'. *Mind*, *109*, 781–823.
21. Shapiro, S. (2006). Vagueness in context. Oxford University Press, Oxford.
22. Smiley, T. (1996). Rejection. *Analysis*, *56*, 1–9.
23. Steinberger, F. (2011). Why conclusions should remain single. *Journal of Philosophical Logic*, *40*, 333–355.
24. Tennant, N. (1999). Negation, absurdity and contrariety. In D. Gabbay, & H. Wansing (Eds.) *What is negation?,* pp. 199–222. Kluwer, Dordrecht.
25. Varzi, A. C. (2007). Supervaluationism and its logics. *Mind*, *116*, 633–676.
26. Williams, J. R. G. (2008). Supervaluationism and logical revisionism. *Journal of Philosophy*, *105*, 192–212.
27. Williamson, T. (1994). *Vagueness.* Routledge, Abingdon-on-Thames.
28. Williamson, T. (2018). Supervaluationism and good reasoning. *Theoria: An International Journal for Theory History and Foundations of Science*, *33*, 521–537.
29. Wright, C. (1976). Language mastery and the sorites paradox. In G. Evans, & J. McDowell (Eds.) *Truth and meaning: essays in semantics,* pp. 223–247. Oxford University Press, Oxford.
30. Yalcin, S. (2007). Epistemic modals. *Mind*, *116*, 983–1026.
31. Zardini, E. (2008). A model of tolerance. *Studia Logica*, *90*, 333–368.
32. Zardini, E. (2013). Higher-order sorites paradox. *Journal of Philosophical Logic*, *42*, 25–48.