Supervaluations Debugged

Josh Dever, Nicholas Asher, and Christopher Pappas

June 3, 2004

Time was that people were tempted to describe modal talk as involving the possession by a simple proposition of a special kind of truth-like property, rather than the possession of simple truth by a complex proposition containing a modal operator. Bradley, for example, inveighs against this temptation in his [1]:

> We must begin by stating an erroneous view. Modality may be supposed to affect the assertion in its formal character, and without regard to that which is asserted. We may take for instance a content S – P, not yet asserted, and may claim for modality the power of affirming this content S – P, unaltered and unqualified, in several ways. S – P, it is supposed, may be asserted, for instance, either simply or problematically, or apodeiktically, and may yet remain throughout S – P: and thus, though the content is unmodified, the assertion is modal.

To make, for example, a distinction between assertoric and apodeictic propositions is to give into this temptation, since it removes a certain class of modal claims from the class of ordinary truth-evaluable claims.

The temptation is one to which we are much less like to submit these days, with modality now treated using object-language operators which enter into the general pattern of assertoric claims. And clearly there are good reasons to resist the temptation, since the inclusion of the modal under the umbrella of the (simply) truth-evaluable makes possible a simple account of inferential connections between the modal and the non-modal and more generally of complex constructions logically relating in a variety of ways the modal and the non-modal. The temptation, however, may be felt more strongly when we remind ourselves that the modal is the semantic correlate of the syntactic phenomenon of mood, and that moods include the interrogative and the imperative in addition to the assertoric and the subjunctive/apodeictic. Questions and commands we still tend to think of as involving a fundamentally different sort of semantic coin, rather than as involving the truth-evaluable consequence of modifying propositions with interrogative and imperative operators (although imbedded questions would seem to create the same pressure toward the object-language inclusion as do the syntactic integration of the assertoric and the apodeictic).

C.I. Lewis' introduction of explicit object-language modal operators marks a large step away from the temptation, and Kripke semantics finish the journey, by showing how to ground the semantics of those operators in the notion of assertoric truth, via the formal tool of truth-at-a-world. When a Kripke semantics is thought of as a tool for a reductive account of the modality, though, it introduces new problems, by creating a demand for an explanation of what worlds are, and how the notion of truth at a world is to be understood. If, however, we dabble with temptation, we can think of modal claims as evaluable in a *sui generis* semantic coin not reducible to assertoric truth, and think of the tools of a Kripke semantics as mere devices for explicating the logical structure of that coin.

The central insight of supervaluation theory is that *vagueness* is a modal phenomenon. This insight entails that a simple notion of truth, when we enter the realm of the vague, needs to be replaced by a notion of modes of truth. Placed within the context of Kripke semantics, the move from a simple to a modalized notion of truth corresponds to the introduction of a structure of points of evaluation, with a base notion of truth-at-a-point and then derived notions of modal truth built off of various configurations of truth-at-a-point. In the particular case of vagueness, the details can be filled out by taking the points of evaluations to be admissible precisifications of the language and the modal truths to be *determinate* truth, *indeterminate* truth, and *supertruth*, but, as with the corresponding application of Kripke semantics to the modality of metaphysical possibility and necessity, care must be taken when deciding how much, and what sort of, weight to rest on the details. One can be a supervaluationist without thinking of vagueness as a linguistic phenomenon – a reflex of a failure of linguistic precision – just as one can endorse modal talk without thinking of necessity as a feature of possible worlds. To do so is to treat the semantic taxonomy of vague language as being fundamentally modal, and hence to refuse to reduce the evaluation of vagueness to the paradigm of assertoric truth.

Supervaluation theory is canonically developed by Kit Fine in his [3], and has subsequently come under attack by Timothy Williamson in [6]. We will argue, though, that the criticisms of supervaluation theory result from failing to take seriously the fundamentally modal character of that theory. That failure is encouraged by obscurities of Kit Fine's presentation of supervaluation theory, which moves back and forth between extensional and intensional perspectives, and consequently introduces some extraneous and damaging elements into the modal analysis of vagueness.

We begin with a presentation of (a slightly simplified version of) Kit Fine's account of supervaluation theory, drawing attention to some intrusions of an extensionalist frame of mind into the development of the theory. After showing how these intrusions make possible Williamson's criticisms of Fine, we set out a modified supervaluation theory which holds true to the modal course. The purely modal supervaluationism is then shown to be immune to Williamson's objections, and to give rise naturally to an account of higher-order vagueness which supports intuitively-desirable "gap principles" without giving rise to contradiction.

1 Fine's Supervaluation Theory

Let a partial model $\mathcal{M}$ be an ordered triple $< D^{\mathcal{M}}, [\![]\!]_+^{\mathcal{M}}, [\![]\!]_-^{\mathcal{M}} >$, where $D^{\mathcal{M}}$ is a domain of quantification, $[\![]\!]_+^{\mathcal{M}}$ is an extension function mapping each n-ary predicate F^n into a subset of D^n, and $[\![]\!]_-^{\mathcal{M}}$ is an anti-extension function mapping each n-ary predicate F^n into a subset of D^n, with the constraint that, for any predicate Π, $[\![\Pi]\!]_+^{\mathcal{M}} \bigcap [\![\Pi]\!]_-^{\mathcal{M}} = \emptyset$.[1] Given an assignment function g, a partial model $\mathcal{M}$ then supports a notion of truth in a model ($\vDash$) and falsity in a model ($\dashv$) with base clauses:

- $\mathcal{M}, g \vDash F^n x_{i_1} \ldots x_{i_n}$ iff $< g(x_{i_1}), \ldots, g(x_{i_n}) > \in [\![F^n]\!]_+^{\mathcal{M}}$
- $\mathcal{M}, g \dashv F^n x_{i_1} \ldots x_{i_n}$ iff $< g(x_{i_1}), \ldots, g(x_{i_n}) > \in [\![F^n]\!]_-^{\mathcal{M}}$,

and recursive clauses of the form:

- $\mathcal{M}, g \vDash \neg\phi$ iff $\mathcal{M}, g \dashv \phi$
- $\mathcal{M}, g \dashv \neg\phi$ iff $\mathcal{M}, g \vDash \phi$
- $\mathcal{M}, g \vDash (\phi \wedge \psi)$ iff $\mathcal{M}, g \dashv \phi$ and $\mathcal{M}, g \vDash \psi$
- $\mathcal{M}, g \dashv (\phi \wedge \psi)$ iff $\mathcal{M}, g \dashv \phi$ or $\mathcal{M}, g \dashv \psi$
- $\mathcal{M}, g \vDash \exists x_i \phi$ iff for some $o \in D^{\mathcal{M}}$, $\mathcal{M}, g[o/x_i] \vDash \phi$
- $\mathcal{M}, g \dashv \exists x_i \phi$ iff for every $o \in D^{\mathcal{M}}$, $\mathcal{M}, g[o/x_i] \dashv \phi$.

Partial model $\mathcal{M}_2$ *extends* partial model $\mathcal{M}_1$, or $\mathcal{M}_1 \geq \mathcal{M}_2$, if:

1. $D^{\mathcal{M}_1} = D^{\mathcal{M}_2}$
2. For all predicates F^n, $[\![F^n]\!]_+^{\mathcal{M}_1} \subseteq [\![F^n]\!]_+^{\mathcal{M}_2}$
3. For all predicates F^n, $[\![F^n]\!]_-^{\mathcal{M}_1} \subseteq [\![F^n]\!]_-^{\mathcal{M}_2}$.

A partial model $\mathcal{M}$ is *complete* if for every predicate F^n, $[\![F^n]\!]_+ \bigcup [\![F^n]\!]_- = D^n$.

A *specification space* is an arbitrary collection of partial models. A *rooted* specification space is a specification space with one partial model identified as privileged (roughly analogous to the actual world of a model for metaphysical modality). A *complete* specification space S satisfies the condition:

- $\forall t \in S \, \exists w \in S \; (w \geq t \;\wedge\; w$ complete).

The language of vagueness will then be given a semantic interpretation using a rooted complete specification space.[2] Call such a space a *Fine space*.

Given a Fine space $\mathcal{F} =< S, r >$, where S is the collection of partial models and r is the root partial model, a sentence ϕ is *supertrue* in $\mathcal{F}$ if ϕ is true (in the sense given

[1]Where no confusion threatens, we omit the superscript model index $\mathcal{M}$.

[2]The requirement of completeness is taken directly from Fine's condition of Completeability. Fine's conditions of Fidelity and Stability fall out from the truth definition given above.

above) at each complete extension of r. Supertruth is thus a modal notion, equivalent to the concatenation of an actuality operator and a necessity operator (taking the partial model extension relation $\geq$ as an accessibility relation). Supervaluation theory then typically takes supertruth as *the* canonical notion of truth, with the result that all classical tautologies are (super)true. A claim is superfalse if it is false at each complete extension of r. (Super)truth-value is thus still gappy, but less gappy than the gappy simple truth-at-r, due to the resolution of tautologies, contradictions, and penumbral truths and falsehoods. [3]

An inference is supervalid if the supertruth of its premises entails the supertruth of its conclusion:

- Γ **superentails** ϕ ($\Gamma \vDash \phi$) if, for every Fine space $\mathcal{F}$, if for every assignment function g and every complete extension w of r in $\mathcal{F}$, $w, g \vDash \Gamma$, then for every assignment function g and every complete extension w of r in $\mathcal{F}$, $w, g \vDash \phi$.

Just as supervaluation theory typically takes supertruth to be the canonical notion of truth, it also typically takes superentailment to be the canonical notion of entailment. By replacing truth in the root specification point with supertruth, the supervaluation theorist can avoid any partial logic.

Superentailment implies that supertruth, should we choose to express it in the object language as a supertruth operator $\mathbb{S}$, has a trivial modal logic: we have both $\phi \vDash \mathbb{S}\phi$ and $\mathbb{S}\phi \vDash \phi$. Of course, this isn't surprising if you want supertruth to act like truth. On the other hand, the equivalence isn't as complete as you might think, if you're thinking along the lines of classical logic. For instance, substitution of $\mathbb{S}\phi$ for ϕ fails in some contexts like the following one: $\phi \leftrightarrow \phi$ is certainly true at every complete specification point (in every model), but $\phi \leftrightarrow \mathbb{S}\phi$ is not, if ϕ is not determined at r. Although the "Tarski biconditional" is not supervalid, a quick inspection verifies that supertruth as an object language operator obeys all the axioms of a **KT5** modal logic.[4]

[3] It should be noted that one of Fine's central arguments for the supertruth theory revolves around its ability to model the various logical relations that hold between vague expressions. Fine labels these relations "penumbral connections" and is particularly concerned with the truths (penumbral truths) that they encode. Such connections hold with respect to predicates and can be either internal or external. For example, a penumbral connection *internal* to the predicate 'bald' would be that if a man with n hairs on his head is bald, then any man with fewer than n cranial hairs is bald, too. Similarly, an *external* penumbral connection might exist between multiple predicates such that "if sociology is to be a science, then so is psychology" (130). Observing such natural connections, Fine introduces as primitive an *admissibility* constraint on the possible models (and, thus, precisifications) that a specification space may contain. He requires that an appropriate specification space consist only of "admissible" models, i.e. models that represent some natural precisification of our current linguistic practices.

[4] Throughout the rest of this paper we discuss various modal logics whose characteristic axioms are listed below. Each modal system represents a certain logic of determinacy whose axioms can be expressed by taking $\mathcal{D}$ as a primitive modal operator, and defining $\mathcal{I}$ as $\neg\mathcal{D}\neg$. The axioms are as follows:

1. **K**: $\mathcal{D}(\phi \rightarrow \psi) \rightarrow (\mathcal{D}\phi \rightarrow \mathcal{D}\psi)$
2. **T**: $\mathcal{D}\phi \rightarrow \phi$
3. **4**: $\mathcal{D}\phi \rightarrow \mathcal{D}\mathcal{D}\phi$
4. **5**: $\mathcal{I}\phi \rightarrow \mathcal{D}\mathcal{I}\phi$

Supervaluation theory accommodates higher-order vagueness through the introduction of a *determinateness* operator, from which a dual *indeterminateness* operator can also be defined. Fine gives two different (and incompatible) definitions of the $\mathcal{D}$ operator:

1. Fine first gives the following clause for $\mathcal{D}$:

   - $w \vDash \mathcal{D}\phi \leftrightarrow r \vDash \phi$.

   $\mathcal{D}$ is thus treated like an actuality operator in metaphysical modality, and would create a **Triv** logic were it not for the influence of supervalidity: the supertruth of ϕ does not entail the supertruth of $\mathcal{D}\phi$ (a tautology, for example, will be supertrue, but may not be true at r), the logic of $\mathcal{D}$ is weaker than **Triv**. The modal axioms **T**, **4**, and **5** continue to hold (but only as inference rules, unless we make the reasonable assumption that every model in a Fine specification space with root r is an extension of **4**), so the resulting logic is **KT5**. Since **KT5** allows no non-trivial iterated modalities, the resulting conception of higher-order vagueness is very thin.

2. Subsequently, with an eye explicitly to the treatment of higher-order vagueness, Fine offers another treatment of the $\mathcal{D}$ operator. The details are both complex and somewhat muddled in Fine's treatment, but the picture amounts to the imposition of a reflexive accessibility relation among the complete specification points and the treatment of $\mathcal{D}$ as a necessity-style operator with respect to that accessibility relation. The result would be the modal logic **KT** (and Fine asserts that it is), but again supertruth complicates the picture. If $\mathcal{D}\phi$ is supertrue, then it is true at every complete specification point. Given the reflexivity of the accessibility relation, it follows that ϕ is true at every complete specification point, and hence supertrue. Thus **T** holds (as an inference principle). But if ϕ is supertrue, then ϕ is true at every complete specification point. Given any complete specification point w, ϕ then holds at every point accessible from w, so $\mathcal{D}\phi$ holds at w. Thus $\mathcal{D}\phi$ is supertrue, so $\mathbf{T_c}$ also holds in the resulting logic, which is therefore **Triv**.

Intertwined with the specification space account of vagueness, Fine runs what he calls the *truth-value account*, on which a vague sentence receives one of some finite number of truth values, with truth values other than **true** and **false** (paradigmatically, **indeterminate**) used to mark the vague territory between paradigm positive and negative instances of a vague property. The truth-value approach is straightforwardly extensional, in its conceptions of truth, of logical consequence (which is simple preservation

5. **B**: $\phi \rightarrow \mathcal{D}\mathcal{I}\phi$
6. $\mathbf{T_c}$: $\phi \rightarrow \mathcal{D}\phi$
7. **Triv**: $\phi \equiv \mathcal{D}\phi$.

Various modal systems will also be referred to, the simplest of which is **K**–the axiom **K** together with the rule of necessitation (i.e. **N:** $\vDash \phi \Rightarrow \vDash \mathcal{D}\phi$). Every other system discussed includes rule **N**, axiom **K**, and some further set of base axioms Λ. Each such system will be referred to as a concatenation of its base axioms, with the exception of **Triv** which shall denote the modal system **K** + the **Triv** axiom. As an example, **KT5** (also known as **S5**) is $\mathbf{K} \cup \Lambda$, where $\Lambda = \{\mathbf{T},\mathbf{5}\}$. The context will determine whether a given name refers to a modal principle or system.

of designated values), and of determinacy (which is taken as an object-language truth-predicate). This extensional line, we will suggest, encourages Fine into an overly extensional view on vagueness , which has ramifications in the specification space account. First, supertruth is treated as a notion of simple truth, rather than as a modal notion. Treating supertruth in this way leads to the second extensional move: treating logical consequence directly in terms of supertruth, rather than in terms of an underlying world-indexed notion of truth. Finally, the specification space account of higher-order vagueness, mirroring a logic perfectly natural on the truth-value account, is designed to produce a logic of **Triv** and support "disquotational" inferences between ϕ and $\mathcal{D}\phi$. In the next section, we trace several problems with which supervaluation theory has been saddled, and in the sequel, show how a return to a more purely intensionalist point of view can eliminate the problems.

2 The Bugs of Supervaluation Theory

Timothy Williamson, in [6], raises two major objections to supervaluation theory: that it fails in its goal of restoring a classical logic to a partial-logic approach to vagueness, and that its notion of (super)truth is not disquotational, and hence not a genuine truth predicate.

The failure of classical logic in supervaluation theory follows from the introduction of the $\mathcal{D}$ operator. Thus, for example, given that the logic of $\mathcal{D}$ is **Triv** (on the final Fine proposal), the inference from ϕ to $\mathcal{D}\phi$ will be valid (i.e., $\phi \vDash \mathcal{D}\phi$). However, the conditional claim $\phi \rightarrow \mathcal{D}\phi$ can fail (of supertruth) in some models. If complete specification points disagree regarding ϕ, then $\mathcal{D}\phi$ will be neither supertrue nor superfalse, so in a complete specification point supporting ϕ, $\phi \rightarrow \mathcal{D}\phi$ will not be true, and thus the conditional will not be supertrue. The logic of supervaluation theory thus lacks a Deduction Theorem: *If* $\Sigma \cup \{\phi\} \vDash \psi$, *then* $\Sigma \vDash \phi \rightarrow \psi$ — a point already noted by Fine. From the failure of the Deduction Theorem, other deviations from classical logic follow: proof by cases, contraposition, and indirect proof all fail.

Williamson argues that supertruth cannot be disquotational:

> Supertruth is not disquotational. If it were, then the supervaluationist would be forced to admit bivalence. Consider any sentence 'A'. By supervaluationist logic, either A or not A. Suppose that supertruth is disquotational. Thus 'A' is supertrue if and only if A and 'Not A' is supertrue if and only if not A. It would then follow, by more supervaluationist logic, either 'A' is supertrue or 'Not A' is supertrue; in the latter case 'A' is superfalse. In order to allow vague sentences in borderline cases to be neither supertrue nor superfalse, the supervaluationist must deny that supertruth is disquotational. ([6], 162)

Suppose, using $\mathbb{S}$ again as an object-language supertruth operator, we formalize the above argument as follows:

1. $A \vee \neg A$	Tautology
2. $A \leftrightarrow \mathbb{S}A$	$\mathbf{T_c}$
3. $\neg A \leftrightarrow \mathbb{S}\neg A$	$\mathbf{T_c}$
4. $\mathbb{S}A \vee \mathbb{S}\neg A$	Disjunctive Syllogism

The first point to notice is that the argument appeals to proof by cases, which fails in the supervaluationist logic. So while the above argument may give another source of that failure, and extend its failure beyond arguments involving $\mathcal{D}$ to arguments involving the other modal operator $\mathbb{S}$, it does not show that $\mathbb{S}$ cannot be disquotational. However, Williamson's argument does act as back-up to the argument from the departures from classical logic – should the failure of the Deduction Theorem be repaired in supervaluationist logic, the argument against disquotationality would be wholly cogent. The second point to note is that the argument appeals to disquotation in the form of explicit object-language biconditionals. While this argument does not show it, the supervaluationist will not accept the biconditional form of the disquotational principles, because when ϕ is gappy, $\mathbb{S}\phi$ will also be gappy, and hence the biconditional will be gappy. What the supervaluationist will accept is inference rules $\phi \vDash \mathbb{S}\phi$ and $\mathbb{S}\phi \vDash \phi$. The extent to which the failure to capture the biconditional formulation of disquotation threatens the sensibility of supertruth as a type of truth is an issue we return to below.

2.1 Problems With Higher-Order Vagueness

In addition to the problems raised by Williamson, supervaluation theory has difficulties with its treatment of higher-order vagueness. The first point has already been noted in passing: the logic of the $\mathcal{D}$ operator is either **KT5** or **Triv**, depending on which Fineian proposal is adopted, and in either case allows for no non-trivial iterations of modal operators beyond either zero or one, and hence allows no real notion of higher-order vagueness. This particular objection is difficult for Williamson to make, since his approach to vagueness also provides only a very limited notion of higher-order vagueness (although he adopts the logic of **KTB**, which allows length-two blocks of modal operators, instead of the length-one of **KT5** and the length-zero of **Triv**.)

Higher-order vagueness can be motivated in the following way. Consider a vague predicate R, and a sequence of objects, consecutive members of which are extremely similar in all R-relevant ways, and gradually moving from an object which is a paradigm instance of R to one which is a paradigm instance of $\neg R$. If the degree of similarity is sufficient, it may fall within the range of vagueness of R, and we may find ourselves unable to place a boundary in the sequence between the objects which are R and those which are not R. We can express this inability via the claim that there are no consecutive members of the sequence, one of which is determinately R and the other of which is determinately not R. Call this the *First-Level Gap Principle*:

- **(G1)** $\forall x(\mathcal{D}Rx \rightarrow \neg\mathcal{D}\neg Rx')$

(where the prime indicates succession in the sequence). Reflection on this state of affairs, however, can lead to a further dissatisfaction. Just as we are unable to place a boundary between the R and the not R, we are also unable to place a boundary between the determinately R and the not determinately R. If the first inability is explained by **(G1)**, then the second should be explained by the claim that there are no consecutive members of the sequence, one of which is determinately determinately R and the other of which is determinately not determinately R. This motivates the *Second-Level Gap Principle*:

- **(G2)** $\forall x(\mathcal{D}\mathcal{D}Rx \to \neg\mathcal{D}\neg\mathcal{D}Rx')$.

Of course, there should be a similar gap between things which are determinately determinately not R and those which are determinately not determinately not R:

- **(G2′)** $\forall x(\mathcal{D}\neg\mathcal{D}\neg Rx \to \neg\mathcal{D}\mathcal{D}\neg Rx')$.

Generalizing, for every n there are two n^{th} level gap principles:

- (**Gn**) $\forall x(\mathcal{D}^n Rx \to \neg\mathcal{D}\neg\mathcal{D}^{n-1}Rx')$
- (**Gn′**) $\forall x(\mathcal{D}\neg\mathcal{D}^{n-1}\neg Rx \to \neg\mathcal{D}^n\neg Rx')$.

Reading $\neg\mathcal{D}\neg$ as an indeterminateness operator $\mathcal{I}$, the gap principles can be thought of collectively as saying that in a sufficiently fine-grained sequence, there is always a region of indeterminateness between two opposing regions of determinateness.

Crispin Wright, in [8], has argued that these gap principles lead to a contradiction. Suppose that the determinateness operator obeys the following rule **DET**:

- **(DET)** If $\Sigma \vDash \phi$, and every atomic sentence of Σ is in the scope of a $\mathcal{D}$ operator, then $\Sigma \vDash \mathcal{D}\phi$.

DET thus amounts to the assumption that the logic of $\mathcal{D}$ obeys both **4** and **5**. We can then reason as follows:[5]

1	1. $\mathcal{D}\mathcal{D}Ra \to \neg\mathcal{D}\neg\mathcal{D}Ra'$	**G2**
2	2. $\mathcal{D}\neg\mathcal{D}Ra'$	A (for conditional proof)
3	3. $\mathcal{D}Ra$	A (for *reductio*)
3	4. $\mathcal{D}\mathcal{D}Ra$	**DET**, 3 (using $\mathcal{D}Ra \vDash \mathcal{D}Ra$)
1,3	5. $\neg\mathcal{D}\neg\mathcal{D}Ra'$	$\to$E,1,4
1,2	6. $\neg\mathcal{D}Ra$	$\neg$I, 2,3,5
1,2	7. $\mathcal{D}\neg\mathcal{D}Ra$	**DET**, 6
1	8. $\mathcal{D}\neg\mathcal{D}Ra' \to \mathcal{D}\neg\mathcal{D}Ra$	$\to$I, 2,7
1	9. $\forall x(\mathcal{D}\neg\mathcal{D}Rx' \to \mathcal{D}\neg\mathcal{D}Rx)$	$\forall$I, 8

Generalizing, we conclude $\forall x(\mathcal{D}\neg\mathcal{D}Rx' \to \mathcal{D}\neg\mathcal{D}Rx)$. It follows that if any member of the sequence is determinately not determinately R, then any earlier member of the series is also determinately not determinately R. Assuming both that anything that is determinately determinately not R is also determinately not determinately R and that

[5]We will appeal directly to instances of the universally quantified gap principles throughout.

anything that is determinately determinately R is also not determinately not determinately not R, it follows that either the sequence contains no object which is determinately determinately R or it contains no object which is determinately determinately not R. Thus higher-order vagueness can play no role.

[2] and [5] have observed that Wright's argument requires the validity of the principle **DET** inside conditional proof, and have argued that supervaluationists should not accept its validity there, for the same reasons that they do not accept the Deduction Theorem for sentences involving the $\mathcal{D}$ operator. Again the conditional structure of the objection to supervaluation theory emerges here: should the initial defect (the non-classical logic) be repaired, a secondary objection (the inconsistency of higher-order vagueness) becomes available. The difficulties with the interaction of conditional proof and **DET** can be avoided, however, by resituating the reasoning in a Gentzen-style sequent calculus:

1. $\mathcal{D}\mathcal{D}Ra \vdash \neg\mathcal{D}\neg\mathcal{D}Ra'$	**G2**
2. $\mathcal{D}Ra \vdash \mathcal{D}\mathcal{D}Ra$	**DET**
3. $\mathcal{D}Ra \vdash \neg\mathcal{D}\neg\mathcal{D}Ra'$	**CUT**, 1,2
4. $\mathcal{D}Ra,\ \mathcal{D}\neg\mathcal{D}Ra' \vdash$	$\neg$-IA, 3
5. $\mathcal{D}\neg\mathcal{D}Ra' \vdash \neg\mathcal{D}Ra$	$\neg$-IS, 4
6. $\mathcal{D}\neg\mathcal{D}Ra' \vdash \mathcal{D}\neg\mathcal{D}Ra$	**DET**

The supervaluation theorist is thus also obligated to reject some of the Gentzen sequent rules – presumably both $\neg$-IA and $\neg$-IS.

[4] presents another formulation of the Wright argument, which avoids both the use of **DET** within conditional proof and the objectional sequent rules by appealing to a stronger inference principle:

- ($\mathcal{D}$-**intro**) If $\Sigma \vDash \phi$, then $\Sigma \vDash \mathcal{D}\phi$.

The rule of $\mathcal{D}$-**intro** requires that determinacy have a modal logic at least as strong as $\mathbf{KT_c}$. Now suppose that we have a series of n objects $a_1, \ldots, a_n$ such that Ra_1 and $\neg Ra_n$. We can then reason as follows:

1	1. Ra_1	A
2	2. $\neg Ra_n$	A
1	3. $\mathcal{D}Ra_1$	$\mathcal{D}$-**intro**, 2
4	4. $\mathcal{D}Ra_1 \rightarrow \neg\mathcal{D}\neg Ra_2$	**G1**
1,4	5. $\neg\mathcal{D}\neg Ra_2$	$\rightarrow$E, 3,4
1,4	6. $\mathcal{D}\neg\mathcal{D}\neg Ra_2$	$\mathcal{D}$-**intro**, 5
7	7. $\mathcal{D}\neg\mathcal{D}\neg Ra_2 \rightarrow \neg\mathcal{D}\mathcal{D}\neg Ra_3$	**G3**
1,4,7	8. $\neg\mathcal{D}\mathcal{D}\neg Ra_3$	$\rightarrow$E, 6,7
1,4,7	9. $\mathcal{D}\neg\mathcal{D}\mathcal{D}\neg Ra_3$	$\mathcal{D}$-**intro**, 8
...	...	...
Γ	3n-1. $\neg\mathcal{D}^{n+1}\neg Ra_n$	$\rightarrow$E, 3n-2,3n-3
2	3n. $\mathcal{D}\neg Ra_n$	$\mathcal{D}$-**intro**, 2
...	...	...
2	4n. $\mathcal{D}^{n+1}\neg Ra_n$	$\mathcal{D}$-**intro**, 4n-1

An explicit contradiction is thus reached, showing that some gap principle must be rejected.

Graff's version of the Wright argument, with its use of the $\mathcal{D}$-**intro** rule, requires a very strong logic of the $\mathcal{D}$ operator. On the **Triv** reading of Fine's formulation of higher-order vagueness, $\mathcal{D}$ will have the required logic, and the argument will go through. However, it is worth noting that a variant of the argument can be formulated which places much weaker constraints on the logic of $\mathcal{D}$. Note first that if we are willing to strengthen our assumptions about the sequence endpoints from Ra_1 and $\neg Ra_n$ to $\mathcal{D}Ra_1$ and $\mathcal{D}\neg Ra_n$, then Wright's rule **DET**, and an **KT5** logic of $\mathcal{D}$, suffices.

Allowing more determinate characterizations of the sequence endpoints then allows for the rule **DET** to be dropped as well. Suppose first that a_1 and a_2 are objects such that a_1 is not just R, but determinately R, and a_2 is not just not-R, but determinately not-R. Then we can derive a contradiction in quick order:

1	1. $\mathcal{D}Ra_1$	A
2	2. $\mathcal{D}Ra_1 \rightarrow \neg\mathcal{D}\neg Ra_2$	**G1**
1,2	3. $\neg\mathcal{D}\neg Ra_2$	$\rightarrow$E, 1,2

The resulting contradiction should, of course, bother no fan of higher-order vagueness or of supervaluation theory. It shows merely that there can't be a two-object sorites sequence with determinate end-points obedient to the gap principles. But such a result comes as no surprise – two objects simply provide insufficient time for the gradual changes the gap principles codify to accumulate to a shift from determinate R-ness to determinate non-R-ness. Given even a third object in the sorites sequence, the argument fails.

Trouble lurks, however, for the three-object sorites sequence. Suppose that we increase the determinateness of the end-points, so that a_1 is determinately determinately R, and a_3 is determinately determinately $\neg R$. Then we will *almost* have a reductio available:

1	1. $\mathcal{D}\mathcal{D}Ra_1$	A
2	2. $\mathcal{D}(\mathcal{D}Ra_1 \to \neg\mathcal{D}\neg Ra_2)$	?
2	3. $\mathcal{D}\mathcal{D}Ra_1 \to \mathcal{D}\neg\mathcal{D}\neg Ra_2$	**K**, 2
1,2	4. $\mathcal{D}\neg\mathcal{D}\neg Ra_2$	→E, 1,3
5	5. $\mathcal{D}\neg\mathcal{D}\neg Ra_2 \to \neg\mathcal{D}\mathcal{D}\neg Ra_3$	**G2**
1,2,5	6. $\neg\mathcal{D}\mathcal{D}\neg Ra_3$	→E, 4,5

In this almost-proof, unlike the previous one for the two-element sorites sequence, we make an assumption about the logic of $\mathcal{D}$ – that it satisfies the **K** rule. This rule, however, is satisfied by any normal modal logic, and is a much weaker assumption than either **DET** or $\mathcal{D}$**-intro**.

The almost-proof can be converted into an actual proof with a justification for the second line. This line is the determinatization of the first gap principle, so if we simply strengthen that gap principle to assert not just that a determinately *R* object is not followed by a determinately not-*R* object, but that it is determinately the case that a determinately *R* object is not followed by a determinately not-*R* object, then the gap in the proof will be plugged and a contradiction ensues.

Again it's not clear that there is immediate cause for concern. Perhaps a three-object sorites sequence is too short to make the transition from determinately determinate *R*-ness to determinately determinate non-*R*-ness without violation of a gap principle; that fact would not in itself seem to pose a threat to the very idea of an adequately gradual transition from determinately determinate *R*-ness to determinately determinate non-*R*-ness. But a general worry begins to emerge. Suppose the sorites sequence contains four objects a_1, a_2, a_3, a_4 such that $\mathcal{D}\mathcal{D}\mathcal{D}Ra_1$ and $\mathcal{D}\mathcal{D}\mathcal{D}\neg Ra_4$. Then we can reason as follows:

1	1. $\mathcal{D}\mathcal{D}\mathcal{D}Ra_1$	A
2	2. $\mathcal{D}\mathcal{D}(\mathcal{D}Ra_1 \to \neg\mathcal{D}\neg Ra_2)$	$?_1$
2	3. $\mathcal{D}\mathcal{D}\mathcal{D}Ra_1 \to \mathcal{D}\mathcal{D}\neg\mathcal{D}\neg Ra_2$	**K**, 2
1,2	4. $\mathcal{D}\mathcal{D}\neg\mathcal{D}\neg Ra_2$	→E, 1,3
5	5. $\mathcal{D}(\mathcal{D}\neg\mathcal{D}\neg Ra_2 \to \neg\mathcal{D}\mathcal{D}\neg Ra_3)$	$?_2$
5	6. $\mathcal{D}\mathcal{D}\neg\mathcal{D}\neg Ra_2 \to \mathcal{D}\neg\mathcal{D}\mathcal{D}\neg Ra_3$	**K**, 5
1,2,5	7. $\mathcal{D}\neg\mathcal{D}\mathcal{D}\neg Ra_3$	→E, 4,6
8	8. $\mathcal{D}\neg\mathcal{D}\mathcal{D}\neg Ra_3 \to \neg\mathcal{D}\mathcal{D}\mathcal{D}\neg Ra_4$	**G3**
1,2,5,8	9. $\neg\mathcal{D}\mathcal{D}\mathcal{D}\neg Ra_4$	→E, 7,8

Here the two questioned steps require the determinate determinatization of the first gap principle and the determinatization of the second gap principle.

Continuing in this vein, we see that the following collection of assumptions is inconsistent:

1. There is a finite-length sorites sequence.
2. The endpoints of the sorities sequence are arbitrarily determinate.

3. The sorites sequence obeys arbitrary determinatizations of all of the gap principles.

4. The $\mathcal{D}$ operator obeys **K**.

Given that there are, for example, color patches that seem to be red in such a way that issues of vagueness in no way impinge on their chromatic categorization, and other color patches that seem to be not red in a similar way, the assumption that arbitrarily determinate sequence endpoints are available seems unproblematic. (Put in Williamson's "margin of error" framework, the point is that the requisite margins may decrease in size as the determinacy hierarchy ascends, converging to a finite cumulative margin small enough to fit inside some range of suitably red samples.) And the motivation for the gap principles appears to yield their determinate truth just as readily as it yields their truth.

The above modification of the Graff/Wright argument shows that the problem extends well beyond supervaluationist treatments of higher-order vagueness (such a conclusion is, of course, in line with Wright's original intention, which was to provide a puzzle for all treatments of higher-order vagueness), since it requires only a minimal logic of $\mathcal{D}$ and not the very strong logic which naturally falls out of supervaluation theory.

Supervaluation theory is then left with the following hierarchy of problems. First, its notion of the determinacy operator gives rise to violations of the Deduction Theorem. The resulting logic is objectionably non-classical. If a major reason to prefer supervaluation theory over a simple multi-valued logic is the preservation of classical logic, this first problem is a pressing one. Second, should the logic be made classical, a simple argument shows the impossibility of a disquotational account of supertruth. Third, supervaluation theory assigns to the $\mathcal{D}$ operator a logic which, if the background logic is classical, wholly trivializes the notion of higher-order vagueness. Fourth, endorsing the gap principles and avoiding the Wright argument, given the supervaluationist acceptance of **DET**, requires a non-classical logic. Fifth, the Graff variant of the Wright argument shows that even with a non-classical logic, a reading of the supervaluationist $\mathcal{D}$ operator which endorses $\mathcal{D}$**-intro** is incoherent, and the present variant of that argument shows that even a much weaker logic of the $\mathcal{D}$ operator is problematic when arbitrary determinizations of the endpoints and gap principles are permitted.

3 Debugging Supervaluation Theory

Things at this point look rather grim for supervaluation theory. How did the simple idea that vagueness is a modal phenomenon lead to such rough waters? Our suggestion is that the difficulties begin when supertruth is given too prominent a position in supervaluation theory. This suggestion can seem perverse – is not supertruth *the* central notion of supervaluation? A supervaluationist overemphasis on the notion of supertruth might seem akin to an accusation that utilitarians overemphasize the notion of maximizing utility.

The difficulty, though, is that supervaluation theory loses sight of the fact that supertruth is a modal notion, akin to actual truth. As such, it represents a *way of being true*, rather than a fundamental notion of truth. Given the Kripkean semantic framework, the fundamental role is played by truth at a world. Depending on one's philosophical attitudes toward the modal theory, one might endorse the fundamentality of truth at a world, or one might (regarding the model theory as more purely heuristic) drop the notion of a fundamental truth notion and simply endorse a plurality of modes of truth.

In either case, the demotion of supertruth immediately calls into question the supervaluationist conception of logical consequence. Logical consequence in modal systems is standardly defined in terms of truth at a world, not in terms of any modally-modified truth. In systems for metaphysical modality, for example, logical consequence is not phrased in terms of preservation of truth at every world. If it were, the accessibility relation would become semantically inert, and the logic of any natural modal operator would become **Triv**.

But this is exactly what is done, and what does happen, in supervaluation theory. Supertruth is, more or less, truth at every specification point, so treating validity as the preservation of supertruth looks at the connection between the global presence of the premise and the global presence of the conclusion. No surprise, then, that the modal logic of $\mathcal{D}$ is naturally impelled toward **Triv**.

An extensionalist frame of mind leads to these supervaluationist troubles. The tendency to treat supertruth as fundamental truth is an extensionalist one – a failure to note the genuine modality of supertruth, perhaps encouraged by Fine's comparisons between the specification space approach and the truth-value approach. We will present a modified, and more purely modal, account of supervaluation theory which restores supertruth to its properly modal position.

3.1 Going Local

Our account begins with a re-evaluation of the basic logical notions that a supervaluationist account introduces. Fine's superentailment relation is based on a global notion of validity, but we note that we could also consider a notion of local entailment on complete extensions that is more familiar from classical logic:

- Γ **locally entails** ϕ ($\Gamma \vDash_l \phi$) if, for every Fine space $\mathcal{F}$, every assignment function g, and every complete extension w of r in $\mathcal{F}$, if $w, g \vDash \Gamma$, then $w, g \vDash \phi$.

In fact, both global and local consequence relations can be given both minimal and maximal readings. On the minimal reading, entailment requires preservation of truth (either globally or locally) at all complete specification points; on the maximal reading, it requires preservation of truth on all specification points, whether complete or not. We thus have four consequence relations $\vDash_g$, $\vDash_G$, $\vDash_l$, and $\vDash_L$. Each will produce its own modal logics for determinacy and for supertruth expressed in the object language, though when we define validity in the usual way ($\models \phi$ holds iff $\varnothing \models \phi$), the two localist conceptions of validity coincide as do the two globalist conceptions.

But the classification of these modal logics is further complicated by the fact that, given the presence of partial models and the failure of the Deduction Theorem, the axiomatic and inferential characterizations of a logic can come apart. There are thus eight cases to consider both for determinacy and for supertruth. For supertruth, we have:

$\mathbb{S}$	Axiomatic Logic	Inferential Logic
Maximal Global ($\models_G$)	**Triv**	$\varnothing$
Minimal Global ($\models_g$)	**Triv**	**KT5**
Maximal Local ($\models_L$)	$\mathbf{T}_c$ **+ D** (without **Triv**)	$\varnothing$
Minimal Local ($\models_l$)	**KT5**	**KT5**

For determinacy, we have:

$\mathcal{D}$	Axiomatic Logic	Inferential Logic
Maximal Global ($\models_G$)	**Triv**	$\varnothing$
Minimal Global ($\models_g$)	**Triv**[6]	**KT**
Maximal Local ($\models_L$)	**KT** (without **Triv**)	$\varnothing$
Minimal Local ($\models_l$)	**KT**	**KT**

If we take as a central feature of a supervaluational account that the classical logical truths be preserved, then we can disregard the *maximal*, localist ($\models_L$) and globalist ($\models_G$) conceptions of consequence, though their logics might interesting in their own right.[7]

[6]On the assumption that we use the second Fineian conception of determinacy set out in section 1. The logic is **Triv** only on the plausible assumption that the accessibility relation never makes incomplete specification points accessible from complete points. One might reject this assumption, and if one does, the inferential rule **Triv** fails. However, without this assumption the rule of necessitation also fails, making moot the whole point of adopting a minimalist conception (preserving classical logic).

[7]On the maximal accounts, the modal logic of supertruth with incomplete specification models is difficult to taxonomize in the classical terminology. Because of the partiality of the models, we must distinguish between:

1. The status of a characteristic modal principle in axiom form (such as the **4** axiom $\Box p \to \Box\Box p$) and the status of that same modal principle taken as a rule of inference (such as $\Box p \models \Box\Box p$), and
2. The question of whether a modal principle is true/valid and the question of whether a modal principle is supertrue/supervalid.

There are thus four cases to consider, two of which collapse:

1. In the simplest case, we note that $\phi \models_G \mathbb{S}\phi$ and $\mathbb{S}\phi \models_G \phi$, so the modal logic of supertruth, considered superinferentially, is simply **Triv**. Such a result is desirable, of course, if supertruth is to behave like a real truth predicate.
2. If we consider the modal logic of truth locally- and globally-inferentially, the situation is more complicated. The principle $\phi \models_L \mathbb{S}\phi$ continues to hold, but the converse principle $\mathbb{S}\phi \models_L \phi$ fails (due to, for example, the tautologies). The resulting modal logic should thus be $\mathbf{T_c}$. However, the **D** principle also holds in this case, since $\mathbb{S}\phi \models_L \neg\mathbb{S}\neg\phi$, and $\mathbf{T_c}$ does not entail **D**. The join of $\mathbf{T_c}$ and **D** is, in the classic case, **Triv**, but the derivation of $\mathbb{S}\phi \models_L \phi$ requires an appeal to contraposition, which is not valid in the partial logic.
3. If we taxonomize the modal logic with maximal consequence relation via its axioms rather than via its inferential principles, then no substantive modal logic results, whether axioms are identified by supertruth or by simple truth (at r). If ϕ is gappy at r, then in some models ϕ will be neither supertrue

We have already seen that supertruth on the minimal globalist conception is inferentially **Triv** but verifies the **KT5** axioms. The minimalist localist conception, on the other hand, fails to verify the inference $\phi \vDash_l \mathbb{S}\phi$, where where $\mathbb{S}$ is an object language supertruth operator. On the other hand, this notion does verify all the **KT5** axioms, as well as the inferential rule, $\mathbb{S}\phi \vDash_l \phi$. This notion seems very well-behaved, and we take it to be the right conception of consequence. In fact, we can collapse the minimal and maximal localist conceptions, if we dispense with the incomplete specification points, which now have little point in our supervaluational approach. Henceforth, let a specification space $\mathcal{S}$ be a non-empty collection of complete models. This allows for a simplification of our localist conception of logical consequence.

- **Consequence Relation$_l$.** $\Gamma \vDash_l \phi$ if, in every specification space $\mathcal{S}$, every assignment function f, and every specification point (or model) w in $\mathcal{S}$, if $\mathcal{S}, w, f \vDash \Gamma$, then $\mathcal{S}, w, f \vDash \phi$.

- **Validity$_l$.** $\vDash_l \phi$ iff $\varnothing \vDash \phi$; or equivalently, ϕ is satisfied at all specification points within all specification spaces relative to all assignment functions.

The localist conception of validity, like the globalist, will coincide with the classical notion for modal-free formulas – formulas containing no occurrences of the $\mathcal{D}$ operator. However, as we will see below, it, unlike the globalist conception, also preserves classical consequence on the modal fragment, and does not produce violations of the Deduction Theorem.[8] Moreover, as we'll see shortly, it's easy to introduce supertruth

nor superfalse, and any conditional putative axiom formed from arbitrary $\mathbb{S}$-prefixings of ϕ will also be gappy.

[8]In a response to McGee and McLaughlin's review of his book *Vagueness*, Williamson presents an argument purporting to show that supervaluational logic, even if it rejects the rule of $\mathcal{D}$-**intro**, still cannot restore classical logic by preserving the deduction theorem. Williamson's argument, directed specifically against McGee and McLaughlin's version of supervaluationism, is couched in terms of a truth predicate, about which he says:

> Traditionally, supervaluationists identified genuine truth and genuine falsity with something like definite disquotational truth and definite disquotational falsity respectively, and validity with the preservation of genuine truth. ([7], 118-119)

Since our version of supervaluation theory does not identify "genuine truth" with definite truth (and, indeed, rejects as overly simplistic the unitary notion of definite truth), we recast the argument here in terms of a determinacy operator. This alteration brings to light again the significance of the move from global to local consequence, and casts further light on the essentiality of a modal conception of supervaluation.

Suppose – as is the case in our system – that we want a determinacy operator $\mathcal{D}$ with a non-trivial logic but with an elimination rule. Then we will reject the rule of $\mathcal{D}$-**intro**. Assuming $\mathcal{D}$ is a normal modal operator, it will also support the principle:

- **(K1)** $\vdash \mathcal{D}(\phi \wedge \psi) \rightarrow (\mathcal{D}\phi \wedge \mathcal{D}\psi)$

Let ϕ be an arbitrary sentence. From $\mathcal{D}$-elimination, we have:

- $\vdash \mathcal{D}(\phi \wedge \neg\mathcal{D}\phi) \rightarrow (\phi \wedge \neg\mathcal{D}\phi)$

From (**K1**), we have:

- $\vdash \mathcal{D}(\phi \wedge \neg\mathcal{D}\phi) \rightarrow (\mathcal{D}\phi \wedge \mathcal{D}\neg\mathcal{D}\phi)$

Thus:

- $\vdash \mathcal{D}(\phi \wedge \neg\mathcal{D}\phi) \rightarrow (\phi \wedge \neg\mathcal{D}\phi \wedge \mathcal{D}\phi \wedge \mathcal{D}\neg\mathcal{D}\phi)$

as a modal notion in this framework.

Let's first sketch the basics. Our supervaluation theory with a $\mathcal{D}$ operator has a straightforward semantics. A model $\mathcal{S}$ for $L_{\mathcal{D}}$ is an 4-tuple $< W, R_{\mathcal{D}}, D, [\![\,]\!] >$, where:

- W is a non-empty set of worlds (or complete specification points)
- $R_{\mathcal{D}}$ is a binary relation on W
- D is a non-empty domain (of individuals)
- $[\![\,]\!]$ is a standard intensional interpretation function for relation symbols, function symbols and constants that for each non-logical constant and world of evaluation gives an appropriate extension.

Truth in a model will be determined in the usual way, with the $\mathcal{D}$ operator acting as a necessity-like operator with respect to the accessibility relation $R_{\mathcal{D}}$.

The axiomatization of the resulting supervaluation theory will then depend on the constraints placed on the accessibility relation $R_{\mathcal{D}}$. There are two ways of approaching the choice of such constraints:

- We can select constraints "from above", by considering the inferential features of determinacy, and selecting accessibility constraints which generate those inferential features. Views on the inferential potential of determinacy will, of course,

---

and:

- $\vdash \neg\mathcal{D}(\phi \wedge \neg\mathcal{D}\phi)$

But suppose an argument is valid just in case it is *determinacy-preserving*:

- **($\mathcal{D}$-$\vdash$)** $\phi \vdash \psi$ iff $\vdash \mathcal{D}\phi \rightarrow \mathcal{D}\psi$

Given $\neg\mathcal{D}(\phi \wedge \neg\mathcal{D}\phi)$, we have:

- $\vdash \mathcal{D}(\phi \wedge \neg\mathcal{D}\phi) \rightarrow \mathcal{D}\psi$

for arbitrary ψ, and hence, via **($\mathcal{D}$-$\vdash$)**:

- $\phi \wedge \neg\mathcal{D}\phi \vdash \psi$

Letting ψ be an arbitrary contradiction, we have:

- $\vdash \neg(\phi \wedge \neg\mathcal{D}\phi)$

and hence:

- $\vdash \phi \rightarrow \mathcal{D}\phi$

In the absence of a rule of $\mathcal{D}$-**intro**, this result contradicts the Deduction Theorem.

The rule **($\mathcal{D}$-$\vdash$)** encodes a version of the global notion of consequence. If one has an extensionalist perspective on supervaluation, and is thinking of determinacy or supertruth as the core notion of truth, then **($\mathcal{D}$-$\vdash$)** will, of course, seem quite natural. But if one moves to thinking of determinacy and supertruth as *modes of truth*, and takes *truth at a world/point* as basic, then **($\mathcal{D}$-$\vdash$)** becomes much less natural, and the local conception of consequence instead becomes preferable. With the local conception, Williamson's argument fails to go through – we still have the invalidity of $\mathcal{D}(\phi \wedge \neg\mathcal{D}\phi)$, but from this invalidity it does not follow that $\phi \wedge \neg\mathcal{D}\phi$ has any special inferential force.

Williamson's argument amounts to the observation that in any normal modal logic whose local consequence relation is at least as strong as **T**, the global consequence relation is **Triv**. We take this fact to be yet another reason to prefer the local over the global consequence relation.

vary. In our view, the only inferential move which is clearly legitimate is that from $\mathcal{D}\phi$ to ϕ, which requires that the accessibility relation by reflexive.

- We can select constraints "from below", by considering the nature of the specification points, thinking about what sort of interesting relation might hold among them, and then working out the structure of that relation. The approach from below has the disadvantage that it requires taking a substantive stand on the nature of the specification points, which we have been trying to suggest is an optional extra in constructing supervaluation theory. It has, however, the advantage that it offers an independent check on the logic of determinacy. The most common supervaluationist view on the specification points is that they represent possible precisifications of the language, with the accessibility relation representing the semantic or pragmatic admissibility of a precisification of a piece of vague language. From this perspective, the most tempting constraints on $R_{\mathcal{D}}$ are:
  - Reflexivity, on the view that the current state of linguistic precision is always acceptable (the inertial view of language).
  - Symmetry, on the grounds that changes in precision can always be withdrawn. Arguably, though, the dynamics of discourse will impose an asymmetry, by privileging admissions over retractions in contextual shifts.
  - Transitivity, on the grounds that if the language can be precisified from point a to point b, and then precisified from point b to point c, then it can be precisified directly from point a to point c.

Since we take the central insight of supervaluation theory to be that vagueness is modal, rather than that vagueness has any particular modal logic, we will be minimally committal about the nature of $R_{\mathcal{D}}$. We will impose reflexivity on $R_{\mathcal{D}}$, because we take the move from $\mathcal{D}\phi$ to ϕ to be partly constitutive of the concept of determinacy. Beyond this, we impose no constraints (but recognize as supervaluational options theories which do impose further constraints). The resulting model theory has a straightforward axiomatization. In addition to the axioms of classical predicate logic and the rules of modus ponens, substitution, and universal instantiation, we have as axioms:

- **K**: $\mathcal{D}(\phi \rightarrow \psi) \rightarrow (\mathcal{D}\phi \rightarrow \mathcal{D}\psi)$
- **T**: $\mathcal{D}\phi \rightarrow \phi$
- the Barcan and converse Barcan formulas: $\forall x\mathcal{D}\phi \leftrightarrow \mathcal{D}\forall x\phi$,

and the rule of necessitation:

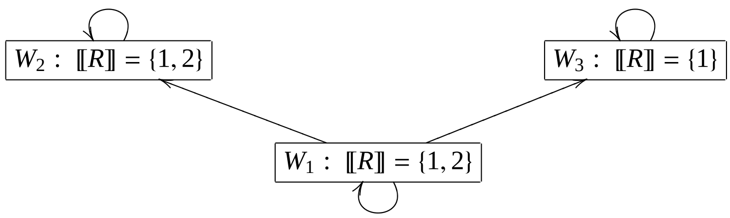
- **N**: $\models_l \phi \Rightarrow \models_l \mathcal{D}\phi$.

The resulting logic does not validate the rule of **$\mathcal{D}$-intro**, because the truth of ϕ at a particular specification point does nothing to guarantee the truth of ϕ at accessible specification points. It also, of course, does not validate the principle $\phi \rightarrow \mathcal{D}\phi$, for similar reasons. Thus the source of violations of the Deduction Theorem in globalist supervaluation theory is removed, and it is straightforward to verify that the Deduction

Theorem holds under local validity. The primary bug in supervaluation theory is thus patched.

With the move to local consequence, a constrained accessibility relation no longer produces a vacuous logic of determinacy, so a genuinely structured notion of higher-order vagueness results. On the minimal constraints we have imposed, the logic of higher-order vagueness is **KT**, which admits of infinitely many non-trivial modal operator prefix blocks, and hence of infinitely many non-redundant levels of higher-order vagueness. The Wright and Graff arguments against higher-order vagueness both fail here, because the $\mathcal{D}$ operator does not obey the requisite **DET** or $\mathcal{D}$**-intro** rules.

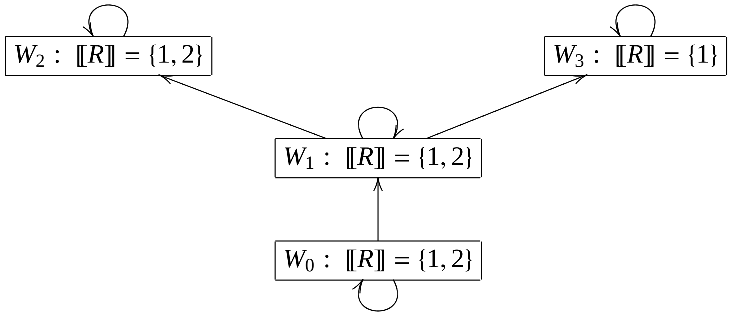
It is, in fact, trivial to give a model in which all of the gap principles are validated. Consider a three-world model with a single unary predicate R and constants a_1, a_2, a_3:



At W_1, we have $\mathcal{D}Ra_1$ and $\mathcal{D}\neg Ra_3$, but neither $\mathcal{D}Ra_2$ nor $\mathcal{D}\neg Ra_2$. Thus **G1** holds. We have $\mathcal{D}\mathcal{D}Ra_1$ but also $\neg\mathcal{D}\neg\mathcal{D}Ra_2$ (due to the determinate truth of Ra_2 at W_2), so **G2** holds. We have $\mathcal{D}\mathcal{D}\neg Ra_3$, but also $\neg\mathcal{D}\neg\mathcal{D}\neg Ra_2$ (due to the determinate falsity of Ra_2 at W_3), so **G2'** holds. Inspection shows that the entire hierarchy of gap principles holds in this model.

The specified model has a transitive accessibility relation, and hence satisfies the **4** axiom, showing that a stronger logic than we have preferred could still validate the gap principles. Note also that the sequence endpoints are not just determinate, but arbitrarily determinate, in the specified model. We thus know, via our modification of the Wright/Graff argument, that the gap principles, while true, cannot be arbitrarily determinately true. In fact, inspection shows that those principles are merely true without being even determinately true – each fails in both W_2 and W_3.

The gap principles can, however, be made determinately true as well. The simplest way of doing so is to prefix another world to W_1 as follows:



The truth of the gap principles at W_1 will then entail the determinate truth of the gap principles at W_0. In this manner, the gap principles can be made determinately-to-the-n true for any n, but cannot be made arbitrarily determinately true.

How significant is it that gap principles cannot be made arbitrarily determinately true? If we think of determinateness in terms of the preservation of aspects of the linguistic practice, it is not clear that it is of any deep significance. The insistence that gap principles be arbitrarily determinately true is in effect an insistence that vagueness is an *ineliminable* feature of the language. But there seems to be no good reason to accept this. If acceptable alterations of the linguistic practice can eventually eliminate vagueness, they will also eliminate gaps and falsify the gap principles. Thus, at any particular point in the linguistic practice, we may have reason to deny that the gap principles are stable given some number n of modificatory steps of the practice.

3.2 More on Supertruth

Thus far our localist reconstruction of supervaluation theory has completely abandoned the notion of supertruth. The price of this abandonment has been a corresponding abandonment of the three-valuedness of the semantic evaluation of vague claims. However, supertruth, properly modally construed, does have a natural home in supervaluation theory as a type of *actuality operator*.

To reintroduce the notion of supertruth within a modal setting, we use another modal operator A that picks out the set of precisifications given by the ground rules for the extension and anti-extension for vague predicates, the first-order vague facts dictated by the language. While supervaluation theory replaces in effect talk of a partial model with talk of a set of its complete extensions or specifications, the modal extension of standard supervaluation theory must distinguish between those extensions that specify the modal free facts, extensions that we'll call base extensions or worlds, and those that specify the modal facts. Our modal operator A will range over those worlds that are supposed to specify the $\mathcal{D}$-free facts. Some but not all of the latter are specification points that represent our actual practice; the rest represent a precisification of our current practices.[9] The worlds that are the second arguments of the accessibility relation represent the precisifications that establish modal facts (those involving $\mathcal{D}$ in their expression) in the base worlds.

Supertruth is just truth in all base worlds. As with all actuality operators, we should expect the logic of A to eliminate iterations of the modality. So not only do the axioms we gave above for $\mathcal{D}$ also hold for A, we should expect the following additional axiom, making the logic for A **KT5**:

- **5**: $\neg A\phi \rightarrow A\neg A\phi$.

A model $\mathcal{S}$ for $\mathcal{L}_{\mathcal{D},A}$ is a 5-tuple $< W, @, R, D, [\![]\!] >$, where,

- W is a non-empty set of worlds (or complete specification points)
- $@ \subseteq W$
- R is a binary, reflexive relation on W
- D is a non-empty domain (of individuals)
- $[\![]\!]$ is a standard intensional interpretation function for relation symbols, function symbols and constants that for each non-logical constant and world of evaluation gives an appropriate extension.

The semantics of formulas of our modal language is standard:

- $\mathcal{S}, w, f \vDash F^n t_{i_1}, \ldots, t_{i_n}$ iff $< [\![t_{i_1}]\!]_g, \ldots, [\![t_{i_n}]\!]_g > \in [\![F^n]\!]$
- $\mathcal{S}, w, f \vDash \neg\phi$ iff $\mathcal{S}, w, g \nvDash \phi$
- $\mathcal{S}, w, f \vDash \phi \wedge \psi$ iff $\mathcal{S}, w, g \vDash \phi$ and $\mathcal{S}, w, g \vDash \psi$
- $\mathcal{S}, w, f \vDash \exists x_i \phi$ iff for some $o \in D$, $\mathcal{S}, w, g[o/x_i] \vDash \phi$
- $\mathcal{S}, w, f \vDash \mathcal{D}\phi$ iff $\forall w'(Rww' \rightarrow \mathcal{S}, w', g \vDash \phi)$
- $\mathcal{S}, w, f \vDash A\phi$ iff $\forall w' \in @, \mathcal{S}, w', g \vDash \phi$.

Note that the actuality operator is not sensitive, as we have defined it, to the particular world or precisification for its evaluation. That is, $\mathcal{S}, w, f \vDash A\phi$ iff $\mathcal{S}, w', f \vDash A\phi$, for any w, w'. We will thus omit the specification of the evaluation world and write $\mathcal{S}, f \vDash A\phi$. With the actuality operator so defined for a vague language, we can now reflect the notion of *supertruth* and *global consequence* within a localist setting. More specifically, we define:

- ϕ is **supertrue** in $\mathcal{S}$ iff $\mathcal{S} \vDash A\phi$[10]
- $\phi_1, \ldots, \phi_n \vDash_g \psi$ iff $A\phi_1, \ldots, A\phi_n \vDash_l A\psi$.

[9] A more fully fleshed out interpretation of the A operator might be epistemic, metaphysical, semantic, or contextual, but we think that this choice is not strictly part of a supervaluation theory.

[10] Henceforth, we suppress mention of the assignment function g, since only closed ϕ are apt for supertruth.

So far we have made no link between what are in fact the possible precisifications corresponding to first-order vagueness and those precisifications that govern higher order vagueness, the alternatives in the range of $R_{\mathcal{D}}$. But since the actuality operator is insensitive to the world of evaluation, the following axioms hold:

- $A\phi \to \mathcal{D}A\phi$
- $\neg A\phi \to \mathcal{D}\neg A\phi$.

These capture at least part of the intuitions behind **$\mathcal{D}$-intro**. We could link $\mathcal{D}$ and A more closely, if we wish, not only to reflect the intuitions behind **$\mathcal{D}$-intro**, but to reflect the fact that the A-related precisifications capture all of the first-order facts about our vague predicates.

- (**AD**) $A\phi \to \mathcal{D}\phi$, for $\mathcal{D}$-free ϕ.

To enforce this principle in the semantics, we require the following constraint on $[\![\cdot]\!]$:

- $\forall w \in W(w \in R[W] \to \exists w' \in @([\![\cdot]\!]|_w = [\![\cdot]\!]|_{w'}))$[11]

The restriction of **AD** to $\mathcal{D}$-free formulas is, as we've seen, essential; the principle fails, for example, when ϕ is a gap principle. With **AD** one might wonder whether we escape from the difficulties that Williamson and others have noticed with $\models_g$ in the presence of $\mathcal{D}$. The Deduction Theorem clearly holds for this **AD** logic. In particular, suppose that if $A\phi$ is true in every model $\mathcal{S}$ for some $\mathcal{D}$-free ϕ, then at every world w of $\mathcal{S}$, $\mathcal{S}, w \models \mathcal{D}\phi$. Then it is immediate that $A\phi \to \mathcal{D}\phi$ is a theorem. Consider now Williamson's argument for the failure of proof by cases. In the localist scenario, the premises of his argument now translate as:

- $A\phi \models_l \mathcal{D}\phi \vee \mathcal{D}\neg\phi$
- $A\neg\phi \models_l \mathcal{D}\phi \vee \mathcal{D}\neg\phi$.

both of which might be true for some $\mathcal{D}$-free ϕ. However, the conclusion that fails to follow from these premises now translates into something that *isn't* the conclusion of a proof by cases from the premises above nor does it hold in our system:

- $A(\phi \vee \neg\phi) \models_l \mathcal{D}\phi \vee \mathcal{D}\neg\phi$.

Instead, we have:

- $A\phi \vee A\neg\phi \models_l \mathcal{D}\phi \vee \mathcal{D}\neg\phi$.

Thus, our introduction of the operator A makes clear where things have gone wrong in the globalist conception. But we have also shown how easy the problems Williamson finds with supervaluationism are to fix.

Finally, how does the properly modalized conception of supervaluation theory respond to Williamson's charge that supertruth is not disquotational? It is helpful to begin by considering why disquotation is a desirable feature of a truth predicate. Why do we want a truth predicate T to support the inferences:

[11] With this semantic constraint, frame completeness of the logic is no longer available. This is a predictable consequence of **(AD)**'s focus specifically on $\mathcal{D}$-free formulas.

- $\phi \vDash T\ulcorner\phi\urcorner$

and:

- $T\ulcorner\phi\urcorner \vDash \phi$.

The implication relation $\vDash$ checks for a coordination in a certain feature – truth – of premise and conclusion. The T predicate reports on the presence of that very feature. Thus when ϕ has the feature which makes it a relevant case for the evaluation of $\phi \vDash T\ulcorner\phi\urcorner$, $T\ulcorner\phi\urcorner$ will, by virtue of reporting on the presence of that feature, also have the requisite feature. When $T\ulcorner\phi\urcorner$ has the feature which makes it a relevant case for the evaluation of $T\ulcorner\phi\urcorner \vDash \phi$, ϕ will, by virtue of $T\ulcorner\phi\urcorner$'s accurate reporting about ϕ, also have that feature.

Crucial to this argument for the disquotational inferences is that the implication relation check for a coordination in *truth* – the property being disquotationally considered. We should not expect *necessary truth* to pass the same disquotational test, because necessary truth is not the feature that the implication relation coordinates. And, in fact, one disquotational direction fails for necessary truth, since $\phi \nvDash \Box\phi$. However, we can restore necessary truth to disquotational status via an implication relation that checks for a coordination in *that very feature*. Define a new consequence relation:

- $\phi \vDash_\Box \psi$ iff if, in any model M and at any world $w \in M$, if ϕ is true at every world w' accessible from w, then ψ is true at every world w' accessible from w.

Given a transitive accessibility relation, we then get the disquotational inferences $\phi \vDash_\Box \Box\phi$ and $\Box\phi \vDash_\Box \phi$.

The property of *actual truth*, on the other hand, will not be disquotational with respect to $\vDash_\Box$, since the truth of $A\phi$ at every world does not entail the truth of ϕ at every world. Actual truth, though, is (as one would expect) disquotational with respect to an inference relation which coordinates actual truth of premise and conclusion.

We should only expect disquotational principles, then, when the property being tested for disquotation is the very same as the property being coordinated by the disquoting inferential relation. With this requirement in mind, the disquotational anomalies of supervaluation theory disappear. While we do not have $\phi \vDash_l \mathbb{S}\phi$ or $\phi \vDash_l A\phi$, we should not expect to have either one, since the inferential relation $\vDash_l$ coordinates *truth at a world*, rather than *supertruth* or *actual truth*.

We could, however, introduce an operator $\mathbb{T}$ which tests for *truth at a world* – $\mathbb{T}\phi$ is true at w iff ϕ is true at w. This operator will then be disquotational with respect to local validity, supporting both $\phi \vDash_l \mathbb{T}\phi$ and $\mathbb{T}\phi \vDash_l \phi$. More generally, if, adopting the modal perspective, we replace homogenous talk of *the* property of truth with heterogenous talk of various *modes of truth*, we need, in thinking about disquotation, to test for the disquotationality of a particular mode of truth using an inferential relation which coordinates that same mode.

4 The Modal Perspective on Supervaluation Theory

The version of supervaluation theory we have given here is intended to validate the central insight that vagueness is a modal phenomenon, by offering two modes of truth – the *actual* and the *determinate* – with which to speak of vague phenomena. The account is quite deliberately underspecified on various points, primarily on the question of what philosophical account of these two modes is to be given. How, if at all, they are related to other familiar modes (epistemic, metaphysical, etc.), and what, if anything, the underlying specification points and accessibility relation are intended to represent are questions which, while certainly of great importance, we don't take to be part of supervaluation theory *per se* to address. Supervaluation theory is thus a genus, rather than a species. Different species within the genus will implement different philosophical programs, and may well give rise to different superlogics of the minimal supervaluationist logic we have implemented here.

As a closing provocative claim, we will assert that, in particular, Williamson's epistemicism is a type of supervaluation theory. Williamson's view adopts the general framework endorsed here, with three particular choices about the nature of the modes:

1. The modality of vagueness is taken to be an epistemic modality, and hence the specification points represent possible epistemic states.

2. The modality of actual truth is taken to be bivalent.

3. The accessibility structure of the specification space is taken to be reflexive and symmetric.

We don't here insist that Williamson is *wrong* in any of these choices, but we do claim that he is doing supervaluation theory, and doing it in a purely optional way. Williamson's choices hang together to some extent – if one starts with the thought that the modality of vagueness is epistemic, one may well end up with the view that the mode of actual truth is bivalent (if one has the further view that there are admissible epistemic states which settle every question) and that the logic is **KTB** (if one accepts, for example, that epistemic indistinguishability is symmetric, and rejects that it is transitive), and if one starts with the thought that the mode of actual truth is bivalent, one may, driven by the epistemic import of classical logic, seek out the epistemic modality as one which naturally gives rise to such bivalence – but they are not the only choices which can be made. Given the *prima facie* surprising consequences of those choices (that, for example, our linguistic practice fully determines classificatory questions about the world), and the range of other choices made available by the general modal supervaluationist perspective, the reasons for following Williamson in his choices are significantly diminished.

References

[1] Frank Bradley. *The Principles of Logic*. William Clowes and Sons, 1883.

[2] Dorothy Edgington. Wright and sainsbury on higher-order vagueness. *Analysis*, 53(4):193–200, 1993.

[3] Kit Fine. Vagueness, truth, and logic. *Synthese*, 30:265–300, 1975.

[4] Delia Graff. Gap principles, penumbral consequence, and infinitely higher-order vagueness. In J.C. Beall, editor, *Liars and Heaps: New Essays on the Semantics of Paradox*. Oxford University Press, 2003.

[5] Jr. Richard Heck. A note on the logic of (higher-order) vagueness. *Analysis*, 53(4):201–2080, 1993.

[6] Timothy Williamson. *Vagueness*. Routledge, 1994.

[7] Timothy Williamson. Reply to mcgee and mclaughlin. *Linguistics and Philosophy*, 27:113–122, 2004.

[8] Crispin Wright. Further reflections on the sorites paradox. *Philosophical Topics*, 15:227–290, 1987.