Chapter 5

Supervaluations

5.1 INCOMPLETE MEANINGS

The problem of vagueness is often conceived as the problem of generalizing a formal theory of meaning applicable only to precise languages to a formal theory of meaning applicable to vague languages too. A Procrustean method of generalization would be to make the vague language precise, then apply the original theory. It is open to several obvious objections. First, a vague language can be made precise in more than one way. Second, the 'can' is only in principle; in practice we cannot make our vague language fully precise in even one way. Third, if a vague language is made precise, its expressions change in meaning, so an accurate semantic description of the precise language is inaccurate as a description of the vague one. These objections seem at first sight to have little in common with each other. However, one line of thought suggests a revision of the Procrustean method that promises to answer them all.

Perhaps the vagueness of a language consists in its capacity in principle to be made precise in more than one way. Not every substitution of precise meanings for vague ones counts as making the language precise, of course. Rather, vague meanings are conceived as incomplete specifications of reference. To make the language precise is to complete these specifications without contradicting anything in their original content.[1] For example, the meaning of 'heap' and the non-linguistic facts are supposed to determine of some things that they are heaps, of others that they are not heaps, and of still others to leave the matter open. The clear cases are those of the first kind, the clear non-cases those of the second, and the borderline cases those of the third. To make 'heap' precise is to assign it a meaning that makes it true of the clear cases, false of the clear non-cases, and either true or false of the borderline cases. Such a sharpening must also respect any systematic constraints built into the meaning of the term. For instance, if x and y are borderline cases of 'heap' whose only relevant difference is that x contains one grain more than y, then the vague meaning of 'heap' already ensures that if y is a heap then so too is x. Thus some sharpenings of 'heap' will be true of both x and y, some will be true of x and not of y, and some will be true of neither, but none will be true of y yet not of x. The original vague meaning of 'heap' is reflected not in any one sharpening but in the class of all its sharpenings.

What is needed is a method that collects many semantic descriptions of the language as made precise in different ways into one semantic description of the original vague language. The first objection above would be answered by considering all ways of making the vague language precise. The second objection would be answered by considering them collectively, without a futile attempt to specify them individually. The third objection would be answered because the incompleteness of a vague meaning is mirrored in the variety of its completions. A method of this kind is the method of *supervaluations*.

5.2 ORIGINS

The method of supervaluations and its application to vagueness seem to have originated in the philosophy of science. It was commonly supposed in the 1950s that scientific vocabulary could be divided into the observational and the theoretical. Observational terms drew their meaning from connections to experience defined by ostension; theoretical terms drew their meaning from connections to observational terms and with each other, as defined by a scientific theory. Vagueness seeped into the observational terms, for it was a commonplace that an ostensive definition using a finite number of particular experiences could not make clear the application of a term for all possible future experience. More important, even if the observational terms were treated as perfectly precise, and a scientific theory were specified to define the connections of the theoretical terms to them and with each other, the meanings of the theoretical terms would not thereby be fixed. Theoretical vocabulary was at best partially defined in terms of observational vocabulary: if it were totally defined in observational terms, it would not after all be theoretical. More than one interpretation of the theoretical terms would respect all the theoretical and ostensive connections.[2] Call any such interpretation *admissible.* On the view just sketched, all admissible interpretations are equally good. Although the plurality of admissible interpretations of the theoretical terms has a different source from the vagueness of the observational ones, both could be taken as forms of indeterminacy in meaning. A natural treatment of theoretical indeterminacy could then be extended to observational vagueness, by a generalization of admissibility to any interpretation meeting all the constraints on meaning.[3]

An admissible interpretation is precise.[4] Each statement in the language is either true on the interpretation or false on it. Call the corresponding assignment of truth-values to statements an admissible *valuation.* Now consider a scientific statement made using theoretical terms. Is it true or false? If each admissible valuation makes it true, then we can say without qualification that it is true. Similarly, if each admissible valuation makes it false, then we can say without qualification that it is false. But if some admissible valuations make it true while others make it false, then neither answer will do, for the interpretations backing one are no better than those backing the other. In this case we seem driven to say that the statement is neither true nor false, even though each admissible valuation makes it either true or false. The *supervaluation* is the assignment of truth to the statements true on all admissible valuations, of falsity to the statements false on all admissible valuations, and of neither to the rest. Few would now endorse the conception of science in which the method of supervaluations originated, for it seems to neglect both the dependence of observation on theory and the dependence of meaning on the actual nature of what is in the environment. Nevertheless, the treatment of indeterminacy it suggested might be the right one for vagueness.

Call a statement *supertrue* if it is true on all admissible interpretations and *superfalse* if it is false on all admissible interpretations. The supervaluationist claims that truth is supertruth and falsity is superfalsity.

The idea of supervaluations, although not the word, is used in Henryk Mehlberg's *The Reach of Science* (1958), and applied to vagueness. He defines a term as vague 'if it can be understood in several ways without being misunderstood'.[5] A statement including vague terms 'is true (or false, as the case may be) if it remains true (or false) under every admissible interpretation of the vague terms it contains'.[6] A non-scientific example of a vague term is 'Toronto', for the spatio-temporal boundaries of its denotation can be admissibly drawn in more than one way. Since 'Toronto is in Canada' is true on each admissible interpretation, it is true. Since 'Toronto is in Europe' is false on each admissible interpretation, it is false. Since 'The number of trees in Toronto is even' is true on some admissible interpretations and not on others, it is neither true nor false.[7]

Much of the interest of the method of supervaluations lies in its treatment of compound statements. Consider, for example, the statement 'The number of trees in Toronto is either odd or even'. It is true on each admissible interpretation, and therefore supertrue. Yet neither 'The number of trees in Toronto is odd' nor 'The number of trees in Toronto is even' is supertrue, for each is false on some admissible interpretations. Since truth is supertruth, according to Mehlberg, a true disjunction may have no true disjunct.[8] He distinguishes between what he calls the logical and meta-logical laws of excluded middle. The logical law of excluded middle is the schema '*A* or not *A*' in the object-language under study; it is now known just as the law of excluded middle. The meta-logical law of excluded middle is the meta-linguistic principle that any statement '*A*' in the object-language is either true or false; it is now known as the principle of bivalence. The supervaluationist denies the meta-logical law but accepts the logical law of excluded middle.[9] The statement '*A* or not *A*' is true on each admissible interpretation, and therefore true, even if '*A*' is true on some admissible valuations and false on others, and therefore neither true nor false, in which case 'Not *A*' is also neither true nor false. In such a case, '*A* or not *A*' is a true disjunction without a true disjunct.

Mehlberg points out that the admissibility of an interpretation cannot be assessed term by term.[10] Consider, for example, the statement 'If Eve is a woman then Eve is an adult', when Eve is a female human on the borderline of adulthood. The statement is clearly true; a woman is an adult female human. Now 'Eve is a woman' is true on some admissible interpretations of ‘woman’ and ‘Eve is an adult’ false on some admissible interpretations of ‘adult’. If the result of combining these interpretations were admissible, then the conditional would have a true antecedent and false consequent on at least one admissible interpretation, and would therefore not be supertrue. The result of combining the interpretations must therefore be inadmissible. One may interpret ‘Eve is a woman’ and ‘Eve is an adult’ as both true, and one may interpret them as both false, but one must not interpret the former as true and the latter as false. The semantic rules for the logical constants achieve a similar effect. On some admissible interpretations ‘Eve is a woman’ is true and ‘Eve is not a woman’ false; on some it is the other way round; but on none do the two statements have the same truth-value.

Mehlberg’s discussion was informal. A more systematic and rigorous treatment of supervaluations (under that name) was given by Bas van Fraassen in the 1960s. However, van Fraassen did not apply the method to the problem of vagueness; he was seeking a semantic treatment of names which lacked a reference and of self-referential sentences such as those involved in the Liar paradox.[11] The application to vagueness was worked out in detail by a number of writers in the 1970s: Michael Dummett, Kit Fine, Hans Kamp, David Lewis, Marian Przelecki.[12]

5.3 LOGIC AND SEMANTICS

A striking feature of supervaluations is the failure of truth-functionality for compound statements. In particular, the standing of a conjunction or disjunction as true, false or neither is not determined by the standings of its conjuncts or disjuncts. Suppose, as before, that ‘Eve is an adult’ and ‘Eve is not an adult’ are neither true nor false. As already noted, ‘Eve is an adult or Eve is not an adult’ is true; but the pleonastic ‘Eve is an adult or Eve is an adult’ is neither true nor false, for it is equivalent simply to ‘Eve is an adult’. Yet the two disjunctions are indistinguishable in terms of the semantic standings of their components; each has two disjuncts that are neither true nor false. Similarly, the two conjunctions ‘Eve is an adult and Eve is not an adult’ and ‘Eve is an adult and Eve is an adult’ are indistinguishable in terms of the semantic standings of their components, each having two conjuncts that are neither true nor false; yet the former conjunction is false and the latter neither true nor false. Contrast a many-valued approach, on which the degree of truth of a conjunction is a function of the degrees of truth of its conjuncts (see also Section 4.14). When ‘Eve is an adult’ has the same degree of truth as its negation, such degree-functionality forces the merely repetitious ‘Eve is an adult and Eve is an adult’ to have as low a degree of truth as the self-contradictory ‘Eve is an adult and Eve is not an adult’. In those circumstances, it also forces ‘Eve is an adult or Eve is an adult’ to have as high a degree of truth as ‘Eve is an adult or Eve is not an adult’. Again, the supervaluation makes ‘Eve is not an adult or Eve is a woman’ true and ‘Eve is an adult and Eve is not a woman’ false; the degree-functional approach assigns them both an intermediate degree of truth.[13]

There is a corresponding difference for conditionals. In the envisaged circumstances, the degree-functional approach cannot distinguish between the obvious ‘Eve is a woman if and only if Eve is an adult’ and the silly ‘Eve is a woman if and only if Eve is not an adult’, since there is no difference in degree of truth on either side of the ‘if and only if’. But each admissible valuation either makes ‘Eve is a woman’ and ‘Eve is an adult’ true and ‘Eve is not an adult’ false or makes ‘Eve is a woman’ and ‘Eve is an adult’ false and ‘Eve is not an adult’ true, so the supervaluation makes ‘Eve is a woman if and only if Eve is an adult’ true and ‘Eve is a woman if and only if Eve is not an adult’ false.

The differences above between the two approaches tell heavily in favour of supervaluations.[14] They are sensitive to intuitively significant distinctions obliterated by degree-functionality. Parallel arguments show the superiority of supervaluations to other forms of many-valued logic. A three-valued logic, for example, based on a classification of sentences as true, false or neither, fares just as badly as one based on a classification according to degree of truth.

What effect have supervaluations on logic? To answer the question, we must first settle on an account of validity appropriate to supervaluations. It might be suggested that if the condition for a sentence to be true is that every admissible valuation makes it true, then the analogous condition for an argument to preserve truth is that every admissible valuation that makes its premises true also makes its conclusion true. Since validity is necessary preservation of truth, an argument is valid just in case necessarily every admissible valuation that makes its premises true also makes its conclusion true. This property may be called *local validity.* The problem for supervaluationists is that supertruth plays no role in the definition of local validity. Yet they identify truth with supertruth; since validity is necessary preservation of truth, they should identify it with necessary preservation of supertruth. That amounts to an alternative definition, on which an argument is valid just in case necessarily if every admissible valuation makes its premises true then every admissible valuation makes its conclusion true, in other words, necessarily if its premises are supertrue then its conclusion is also supertrue. The latter property may be called *global validity.* An admissible valuation on which '*A*' is true and '*B*' false automatically makes the argument from '*A*' to '*B*' not locally valid, but the argument might still be globally valid, for if '*A*' is false on another admissible valuation, this is not a case in which '*A*' but not '*B*' is supertrue. It is obvious that a locally valid argument is also globally valid. In many languages, the converse also holds: any globally valid argument is also locally valid. Indeed, in any language the converse holds for an argument without premises. But examples will be given later of other arguments that are globally but not locally valid. For the reason given above, we may work with the assumption that supervaluationists identify validity with global validity. From time to time the consequences of the alternative identification with local validity will also be noted.[15]

In simple cases, global validity coincides with classical validity, as it does with local validity. Consider, for example, any argument containing no constants other than negation, conjunction, disjunction, material implication, identity and the universal and existential quantifiers. The premises and conclusion may contain propositional or predicate variables, which can have any interpretation appropriate to their syntactic category. If the argument is classically valid, then since admissible valuations are classical, any admissible valuation that makes the premises true also makes the conclusion true; thus the argument is locally valid, and therefore globally valid. Conversely, it can be shown that if it is not classically valid, then there is an assignment of precise values to its variables on which its premises are true and its conclusion false, so it is neither locally nor globally valid. Within this language, local, global and classical validity are equivalent. In particular, a single formula as the conclusion of an argument without premises is globally valid if and only if it is classically valid, whence the global validity of the law of excluded middle. Thus supervaluationism seems to inherit the power of classical logic.

Supervaluations validate classical predicate logic, but they also enable it to be extended. They make room for a new operator 'definitely' to express supertruth in the object-language: 'Definitely *A*' is true if and only if '*A*' is supertrue. For example, to say that something is neither definitely a heap nor definitely not a heap is to say that it is a borderline case. 'Definitely' can be given a formal semantics very like the possible worlds semantics for the modal operator 'necessarily'. For simplicity, the analogy will first be developed by reference to the simple modal system S5. Account is taken of further complications in Section 5.6. The aim of the formal semantics is to define in mathematical terms a set of models such that an argument is valid if and only if it preserves truth in every model in the set, for that will provide us with a precise standard of validity.

A model for S5 is a *structure* containing a number of objects, which may be thought of as *possible worlds*. Each world in such a structure is associated with a valuation: sentences are true or false at worlds. 'Not *A*' is true at a world if and only if '*A*' is not true at that world, '*A* and *B*' is true at a world if and only if '*A*' is true at that world and '*B*' is true at that world, and so on. In contrast, the truth-value of 'Necessarily *A*' at a world in a structure depends on its truth-values at all the worlds in the structure, not just that one; 'Necessarily *A*' is true at a world in a structure if and only if *A* is true at every world in the structure. By analogy, a model for the language with 'definitely' is a *space* containing a number of *points*, which may be thought of as admissible interpretations. Each point in such a space is associated with a valuation: sentences are true or false at points. 'Not *A*' is true at a point if and only if '*A*' is not true at that point, '*A* and *B*' is true at a point if and only if '*A*' is true at that point and '*B*' is true at that point, and so on. The truth-value of 'Definitely *A*' at a point in a space depends on its truth-values at all the points in the space, not just that one; 'Definitely *A*' is true at a point in a space if and only if '*A*' is true at every point in the space.

A formula in the language of 'necessarily' is valid in the S5 semantics if and only if it is true at every world in every structure. Thus every instance of the axiom schema 'If necessarily if *A* then *B*, and necessarily *A*, then necessarily *B*' is valid in S5, for when 'If *A* then *B*' and '*A*' are both true at every world in a structure, so too is '*B*'. Moreover, if '*A*' is valid in S5, then it is true at every world in any structure, so 'Necessarily *A*' is true at every world in any structure, so 'Necessarily *A*' is valid in S5 too (the rule of necessitation). Quite generally, the logical consequences of necessary truths are themselves necessary truths. 'If necessarily *A* then *A*' (known as the T schema) is also valid in S5, for if '*A*' is true at every world in a structure, then it is true at any world in that structure. What necessarily holds, holds. The distinctive S5 axiom schema is 'If not necessarily *A* then necessarily not necessarily *A*'; it is valid in S5 because the semantics makes the addition of 'necessarily' to 'Not necessarily A' vacuous. For similar reasons, 'If necessarily *A* then necessarily necessarily *A*' (known as the S4 schema) is also valid in S5. According to S5, it cannot be contingent whether something is necessary.

By analogy, a formula in the language of 'definitely' is valid if and only if it is true at every point in every space. Thus 'If definitely if *A* then *B*, and definitely *A*, then definitely *B*' is valid. If '*A*' is valid, then 'Definitely *A*' is valid too. Quite generally, the logical consequences of definite truths are themselves definite truths. The T schema 'If definitely *A* then *A*' is valid. What definitely holds, holds. The S5 schema, 'If not definitely *A* then definitely not definitely *A*' is valid, as is the S4 schema 'If definitely *A* then definitely definitely *A*'. On this semantics, it cannot be indefinite whether something is definite.

Validity has so far been considered only for single formulas. However, the more important notion is of validity for arguments. Here the analogy between 'definitely' and 'necessarily' begins to break down. An argument is valid in S5 if and only if in any structure, if its premises are true at a given world, then so is its conclusion. For 'definitely', one might analogously define an argument as valid if and only if in any space, if its premises are true at a given point, then so is its conclusion. However, this is the formal analogue of *local* validity. It was argued above that supervaluationists should identify validity with *global* validity. They should therefore use its formal analogue: an argument is (globally) valid just in case in any space, if the premises are true at *every* point, then so is the conclusion.

For single formulas, local and global validity coincide, so the analogy between 'definitely' and 'necessarily' remains. The disanalogies emerge for arguments with at least one premise. If '*A*' is an atomic formula, the inference from '*A*' to 'Necessarily *A*' is not valid in S5, for '*A*' can be true at one world in a structure and false at another world in the same structure, so that 'Necessarily *A*' is false at the former world, and validity in S5 requires each world to preserve truth. The converse inference from 'Necessarily *A*' to '*A*' is of course valid in S5. For 'definitely', in contrast, global validity merely requires each space to preserve supertruth (truth at every point), so the inference from '*A*' to 'Definitely *A*' is globally (but not locally) valid, for if '*A*' is true at every point in a space, then so is 'Definitely *A*'. The converse inference from 'Definitely *A*' to '*A*' is of course globally valid.

Since '*A*' and 'Definitely *A*' are interderivable, one might expect 'definitely' to be a redundant operator. It is not, however, for 'If *p* then definitely *p*' is not globally valid, where '*p*' is an atomic formula. If '*p*' is true at some but not all points in a space, then the conditional has a true antecedent and false consequent.[16] For similar reasons, the inference from 'Not definitely *p*' to 'Not *p*' is not globally valid. Such examples can be shown to involve breakdowns of the classical rules of contraposition, conditional proof, argument by cases and *reductio ad absurdum* in the supervaluationist logic of 'definitely'. This is in a sense a violation of classical propositional logic, but at the level of inference rules permitting transitions from arguments to arguments rather than from formulas to formulas. The cases are as follows.

(a) *Contraposition* In classical logic, if one can validly argue from '*A*' and auxiliary premises to '*B*', then one can validly argue from 'Not *B*' and the auxiliary premises to 'Not *A*'. Contraposition does not hold in the supervaluationist logic, for although the inference from '*p*' to 'Definitely *p*' is globally valid, the inference from 'Not definitely *p*' to 'Not *p*' is not globally valid.[17]

(b) *Conditional proof* (*the deduction theorem*) This is the standard way of reaching conditional conclusions. In classical logic, if one can validly argue from '*A*' and auxiliary premises to '*B*', then one can validly argue from the auxiliary premises alone to 'If *A* then *B*'. Conditional proof does not hold in the supervaluationist logic, for although the inference from '*p*' to 'Definitely *p*' is globally valid, the unpremised conclusion 'If *p* then definitely *p*' is not globally valid.[18]

(c) *Argument by cases* (*or-elimination*) This is the standard way of using disjunctive premises. In classical logic, if one can validly argue from '*A*' and auxiliary premises to '*C*', and from '*B*' and auxiliary premises to '*C*', then one can validly argue from '*A* or *B*' and the combined auxiliary premises to '*C*'. Argument by cases does not hold in the supervaluationist logic, for although the inference from '*p*' to 'Definitely *p* or definitely not *p*' is globally valid, as is that from 'Not *p*' to the same conclusion, the inference from '*p* or not *p*' to 'Definitely *p* or definitely not *p*' is not globally valid.

(d) *Reductio ad absurdum* This is the standard way of reaching negative conclusions. In classical logic, if one can validly argue from '*A*' and auxiliary premises to '*B*', and from '*A*' and auxiliary premises to 'Not *B*', then one can validly argue from the combined auxiliary premises to 'Not *A*'. *Reductio ad absurdum* does not hold in the supervaluationist logic, for the inference from '*p* and not definitely *p*' to 'Definitely *p*' is globally valid, as is that from the same premise to 'Not definitely *p*', but the unpremised conclusion 'Not: *p* and not definitely *p*' is not globally valid (it is equivalent to 'If *p* then definitely *p*').[19]

Conditional proof, argument by cases and *reductio ad absurdum* play a vital role in systems of natural deduction, the formal systems closest to our informal deductions. They are the rules by which premises are discharged, i.e. by which categorical conclusions can be drawn on the basis of hypothetical reasoning. Contraposition is another very natural deductive move. Thus supervaluations invalidate our natural mode of deductive thinking. The examples so far have all involved the 'definitely' operator. If we had to exercise caution only when using this special operator, then our deductive style might not be very much cramped. However, supervaluationists have naturally tried to use their semantic apparatus to explain other locutions. If their attempts succeed, our language will be riddled with counterexamples to the four rules (see Section 5.5). In order to restore classical logic, supervaluationists might switch to the alternative definition of validity as local validity. That would restore classical logic at the expense of endangering the identification of truth with supertruth, for validity would no longer be identified with the preservation of supertruth. The gravity of the danger will emerge in Section 5.7.

5.4 THE ELUSIVENESS OF SUPERTRUTH

According to supervaluationism, 'p or q' is sometimes true when no answer to the question 'Which?' is true. For similar reasons, 'Something is F' is sometimes true when no answer to the question 'Which thing is F?' is true. In this sense supertruth is elusive.

The most dramatic examples occur in sorites paradoxes. Consider the Heap. Any admissible valuation has a cut-off number k for 'heap', more than one and less than ten thousand. On the valuation, 'n grains make a heap' is true if n is more than k and false otherwise; in particular, '$k + 1$ grains make a heap' is true and 'k grains make a heap' false, so '$k + 1$ grains make a heap and k grains do not make a heap' is true, so 'For some n, $n + 1$ grains make a heap and n grains do not make a heap' is true. Since the existential generalization is true on each admissible valuation, it is supertrue. Yet no answer to the question 'For which n do $n + 1$ grains make a heap and n grains not make a heap?' is supertrue, for not all admissible valuations have the same cut-off number. The supervaluational response to the sorites argument from 'For all n, if $n + 1$ grains make a heap then n grains make a heap' and 'Ten thousand grains make a heap' to 'One grain makes a heap' is now clear. The argument is classically valid, and therefore globally (and locally) valid. The minor premise 'Ten thousand grains make a heap' is supertrue, and the conclusion 'One grain makes a heap' superfalse. However, each admissible valuation has a counterexample to the major premise, although it is not the same in each case; thus the major premise is superfalse. In short, the argument is valid but unsound.[20]

The supervaluational treatment of the sorites argument is formally elegant. The question is whether it defuses the intuitive backing for the major premise. Many people have found the major premise plausible just because it seemed to them that there could not be a number n such that $n + 1$ grains make a heap and n do not. Supervaluationism makes the very claim that they find incredible. Nor should the supervaluationist say that the claim does not mean what they think it means. The point of the enterprise is to give semantic descriptions of vague sentences as we actually use them. If supervaluationism delivers a meaning for the existential claim other than its ordinary one, the enterprise fails.

The supervaluationist must insist that the sentence ‘For some *n*, *n* + 1 grains make a heap and n grains do not make a heap’ is true in its ordinary sense, and use the apparatus of supervaluations to explain away appearances to the contrary. Usually, a true existential generalization has a true instance; we note that no sentence of the form ‘*n* + 1 grains make a heap and n grains do not make a heap’ is true, and naturally tend to assume that the corresponding existential generalization is not true either. In effect, the explanation is that we ignore vagueness, making semantic assumptions appropriate only if ‘heap’ were not vague. The trouble with the explanation is that it assumes that we do not ignore vagueness at a different point. It is precisely because we have noticed the vagueness of ‘heap’ that we doubt that anything of the form ‘*n* + 1 grains make a heap and *n* grains do not make a heap’ can be true. Perhaps we are so confused that we notice the vagueness and ignore its consequences. However, there is at least a suspicion of a mismatch between supertruth and truth in the ordinary sense. This suspicion will be confirmed in Section 5.7.[21]

5.5 SUPERVALUATIONAL DEGREES OF TRUTH

The idea of degrees of truth tends to be associated with the assumption that the degree of truth of a complex sentence is a function of the degrees of truth of its components, and in particular with many-valued logic. However, Lewis and Kamp showed that it can be understood in terms of supervaluations too.[22] As a first try, suppose that ‘*B*’ is true on every admissible interpretation on which ‘*A*’ is true; in other words, the material conditional ‘If *A* then *B*’ is supertrue. One might then say that ‘*B*’ is *at least as true as* ‘*A*’. If ‘*B*’ is at least as true as ‘*A*’ but ‘*A*’ is not at least as true as ‘*B*’, then one might say that ‘*B*’ is *truer than* ‘*A*’. For example, different admissible interpretations will set different precise standards for counting as ‘young’. However, if Adam was born before Eve, then she will count as ‘young’ by any standard by which he does.

On any admissible interpretation on which 'Adam is young' is true, 'Eve is young' is true too; thus 'Eve is young' is at least as true as 'Adam is young'. Correspondingly, the conditional 'If Adam is young then Eve is young' is supertrue. Assume that there is a reasonable standard by which Eve counts as 'young' and Adam does not. Then 'Adam is young' is not at least as true as 'Eve is young', so the latter is truer than the former.

That first attempt is rather crude. For example, one might sometimes wish to say that 'Eve is young' and 'Eve is not young' are equally true, in the sense that each is at least as true as the other. According to the definitions above, that is impossible, for '*A*' and '*B*' are as true as each other if and only if they are true on exactly the same admissible interpretations, which contradictories never are (given that there is at least one admissible interpretation). Nevertheless, 'Eve is young' and 'Eve is not young' might be thought to be equally true in the sense that the set of admissible interpretations on which the former is true and the set on which the latter is are equally 'large'. This idea presupposes a measure of the size of sets of admissible interpretations. Such a measure might boldly, or rashly, be postulated.

Once a notion of 'truer than' is in place, it can be used to formulate a semantic treatment of comparative adjectives. The guiding principle is that '*a* is *F*er than *b*' is true if and only if '*a* is *F*' is truer than '*b* is *F*'. For example, 'Eve is younger than Adam' is true if and only if 'Eve is young' is truer than 'Adam is young'. More precisely, '*a* is *F*er [more *F*] than *b*' is true on an admissible valuation if and only if '*a* is *F*' is truer than '*b* is *F*'. A corresponding account can be given of 'at least as'. '*a* is at least as *F* as *b*' is true on an admissible valuation if and only if '*a* is *F*' is at least as true as '*b* is *F*'.[23] Lewis and Kamp extended the treatment of comparatives to modifiers such as 'rather' and 'in a sense', as in 'rather clever' and 'clever in a sense'.[24]

Semantic treatments of the kind above face a number of related problems that seem to be caused by the use of a fixed class of admissible interpretations. They may be compared with the similar problems faced by degree-theoretic treatments of such constructions within a framework of many-valued logic (Section 4.11).

(a) Since 'truer than' is defined in terms of *all* admissible valuations, the truth value of '*a* is *F*er than *b*' should be stable from one admissible valuation to another. Thus the semantics predicts that comparatives and related terms should be absolutely precise. Now 'taller' does indeed seem more precise than 'tall'. But it does not seem perfectly precise; stoops and curly scalps may produce borderline cases even for it. A comparative such as 'more intelligent' is notably vague.

(b) If 'truer than' is defined in terms of admissible valuations, '*A*' is not truer than '*B*' when '*B*' is true on every admissible valuation; so if '*A*' is truer than '*B*', '*B*' is not (super)true. Since the truth of 'David is braver than Saul' requires 'David is brave' to be truer than 'Saul is brave', it is incompatible with the truth of 'Saul is brave'. Thus 'Saul is brave, but David is braver than Saul' cannot be true. That is absurd. The brave are not all equally brave.[25]

(c) Consider 'acute' as an adjective of angles. It is precise, for '*a* is acute' is true if *a* is less than a right angle, and false otherwise. 'An angle of 60° is acute' is true, and therefore true on every admissible valuation. Nevertheless, an angle of 30° is more acute than an angle of 60°.

(d) One would expect that a good semantics of comparatives would extend to a modifier such as 'very'. Now if 'very' is treated by means of admissible valuations, it should be sufficient for the truth of '*a* is very *F*' that '*a* is *F*' is true on every admissible valuation. But then 'That is definitely dark blue' should entail 'That is very dark blue', which it does not.

The moral of (a) is presumably that 'admissible' is itself a vague notion. The resolution of the problem therefore depends on an adequate supervaluationist treatment of higher-order vagueness. Such a treatment may also provide the means to resolve (b)–(d).[26] More generally, one of the chief challenges to supervaluationism is to make proper room for higher-order vagueness. We must therefore give a supervaluationist account of the phenomenon, before briefly returning to (a)–(d).

5.6 SUPERVALUATIONS AND HIGHER-ORDER VAGUENESS

A supervaluation divides the sentences of a language into three classes: the true, the false, and the neither true nor false. The comprehensiveness of the classification is plausible enough in some applications of the method. Consider, for example, the view that the future is open except in so far as it is determined by laws of nature and the present state of the universe. On this view, a future tense statement is now true if it is true of each possible future, false if it is false of each possible future, and neither true nor false otherwise. Since each possible future corresponds to a bivalent valuation, the method of supervaluations is appropriate. Whatever its merits, this view makes the threefold division quite natural. Vagueness is a different matter. If it is hopeless to look for the first red shade in a sorites series from orange to red, it is equally hopeless to look for the first shade which can truly be called 'red' (try). The idea that our rough-and-ready use of vague terms does not determine hidden boundaries tells just as much against a pair of hidden second-order boundaries between the true and the neither true nor false and between the latter and the false as it does against one hidden first-order boundary. In supervaluationist terms, the admissibility of a valuation is itself a vague notion.[27]

Second-order vagueness shows itself in an object-language with a 'definitely' operator. If vagueness were only first order, 'Definitely *A*' would be precise, so 'Either definitely definitely *A* or definitely not definitely *A*' would hold. But 'Definitely *A*' has borderline cases, for otherwise there would be sharp second-order boundaries. Since borderline cases for '*A*' are counterexamples to the schema 'Either definitely *A* or definitely not *A*', borderline cases for 'Definitely *A*' are counterexamples to the schema 'Either definitely definitely *A* or definitely not definitely *A*'. That schema entails the S5 schema 'If not definitely *A* then definitely not definitely *A*', for 'If not definitely *A* then not definitely definitely *A*' is uncontroversial.[28] It follows that the simple form of supervaluationist semantics described in Section 5.3 is inadequate, for it validates the S5 schema.

Third-order vagueness is equally possible. If it were not, everything would be either definitely definitely definitely red or definitely not definitely definitely red: but where would the break come? The point extends to any order. If (n + 1)th-order vagueness were not possible, everything would be either definitely definitely^n red or definitely not definitely^n red. The idea that our rough-and-ready use of a vague term does not determine hidden boundaries tells against a hidden boundary between the extension of the first disjunct and the extension of the second. Even if high orders of vagueness are somehow ruled out by empirical factors, that would not entitle the logician to treat them as impossible. Thus supervaluationist logic should admit at least all finite orders of vagueness.[29]

Supervaluationists often regard admissibility as consistency with the semantic rules of the language. If the rules decide a case, then an admissible interpretation decides it in the same way; it may decide a case when they do not. Since consistency is a matter of logic, admissibility looks as though it should be a precise concept.[30] Higher-order vagueness shows this picture to be misleading. A vague meaning is not like a partial definition in mathematics, formulated in precise terms but not covering all cases. If a vague term is governed by semantic rules, then they are formulated in equally vague terms. Moreover, it is not plausible that the limits of higher-order vagueness are laid down by a hierarchy of higher-order semantic rules. When admissibility is not pictured as consistency with semantic rules, supervaluationism becomes a less inviting approach. Nevertheless, it can adapt to higher-order vagueness. Admissibility might be conceived as a matter of *reasonableness.* An interpretation is reasonable if it does not license misuses of the language (from the standpoint of an ordinary understanding of it).[31] The concept of a misuse is obviously and essentially vague.

The formal semantics for 'definitely' in Section 5.3 validated the S5 axiom. A new formal semantics is therefore required to make room for higher-order vagueness. A similar problem arises in modal logic with the semantics for systems weaker than S5. There, the informal idea is that the possibility of a world is itself a contingent matter. The formal trick is to introduce a relation of accessibility between worlds in a structure; 'Necessarily A' is true at a world w if and only if 'A' is true at every world accessible from w. Analogously, one can introduce a relation of admitting between points in a space; 'Definitely A' is true at a point s if and only if 'A' is true at every point admitted by s. The informal idea is that the admissibility of an interpretation is itself a matter for interpretation. Each interpretation makes its own ruling as to which interpretations are admissible. Formally, a point determines both a bivalent valuation and a set of admitted points. Every point should admit itself; were admitting not a reflexive relation, 'A' might be true at every point admitted by a point s yet not at s itself, in which case 'If definitely A then A' would be false at the point. An interpretation should regard at least itself as reasonable.

In order to accommodate higher-order vagueness, admitting is allowed to be non-transitive. Informally, an interpretation might admit just those interpretations that are reasonable by its lights, because they do not differ from it by too much. If you regard me as reasonable and I regard him as reasonable, you may not regard him as reasonable, for the difference between you and him may be too much even if neither the difference between you and me nor that between me and him is too much. Suppose that a point s admits a point t, which admits a point u, but s does not admit u, and 'A' is true at every point admitted by s but not at u. Then 'Definitely A' is true at s but not at t, so 'Definitely definitely A' is false at s, and so the S4 principle 'If definitely A then definitely definitely A' is false at s. By similar reasoning, the S5 principle 'If not definitely A then definitely not definitely A' also fails. Indeed, in the absence of special restrictions on admitting, the valid formulas are just those corresponding to the theorems of the weak modal logic T, which are just those derivable from the principles other than the S4 and S5 schemata listed in Section 5.3 as valid for 'definitely'.[32]

The interpretation dependence of admissibility is exactly analogous to the contingency of possibility in modal logics weaker than S5. What supervaluationism adds is a conception of admissibility. The conception needs to be worked out with some care. Interpretations specify lists of admissible interpretations, including themselves; such interpretations might be suspected of vicious circularity. Fortunately, the circularity can be avoided. A 0-level interpretation makes precise those expressions of the language that, unlike 'definitely', do not have to do with admissibility. A 1-level interpretation specifies which 0-level interpretations are admissible. More generally, an $(i + 1)$-level interpretation specifies which i-level interpretations are admissible. An ω-level interpretation is an infinite sequence $s_0 s_1 s_2 \ldots$, where each s_i is an i-level interpretation and each s_{i+1} specifies that s_i is admissible. An ω-level interpretation $s_0 s_1 s_2 \ldots$ admits an ω-level interpretation $t_0 t_1 t_2 \ldots$ if and only if each s_{i+1} admits t_i; in non-relational terms, $t_0 t_1 t_2 \ldots$ is admissible if and only if each t_i is admissible. A point in a space is the formal analogue of an ω-level interpretation. Thus every point admits itself, but the definitions can be shown to impose no further constraint on the structure of admitting. Further constraints could be added, but they will not be considered here.

An objection might be raised to the foregoing account of higher-order vagueness. Define 'Definitely* *A*' to mean the infinite conjunction: *A* and definitely *A* and definitely definitely *A* and . . . . The definition guarantees that if definitely* *A* then definitely definitely* *A* and indeed definitely* definitely* *A*. In terms of the formal semantics, let a point *s* admit* a point *t* if and only if either *s* admits *t*, or *s* admits a point that admits *t*, or *s* admits a point that admits a point that admits *t*, or . . . . 'Definitely* *A*' is true at a point *s* if and only if '*A*' is true at every point that *s* admits*. In technical terms, admitting* is the ancestral of admitting; it is automatically transitive, even though admitting is not. The supervaluationist approach can now be applied in terms of admissibility* rather than admissibility. Since the strict notion 'definitely*' obeys an S4 axiom, higher-order vagueness disappears. It turns out to be a surface phenomenon, reflecting our use of an unnecessarily lax standard of definiteness. According to the objection, the supervaluationist cannot recognize higher-order vagueness as a deep phenomenon. The point might be used against supervaluationism; it might be used against the claim that higher-order vagueness is a deep phenomenon.[33]

The supervaluationist may insist that even 'definitely*' is vague. Its vagueness is not manifest in its failure to obey some principle stated in its own terms, but that is just to say that it cannot be used to measure its own vagueness. It is like a cloud said to have an exact length because it is exactly as long as itself. A new operator 'definitely!' is needed to express the vagueness of 'definitely*' in the failure of the principle 'If definitely* *A* then definitely! definitely* *A*'. The vagueness of 'definitely*' corresponds to a vagueness in the meta-language for the original language of 'definitely' on which there had been no need to reflect before 'definitely*' was introduced. The meta-language for the new language of 'definitely!' may in turn harbour vagueness whose expression will require yet another operator 'definitely!!'. The process may have no natural end.

An alternative supervaluationist reply is that 'definitely*' is precise, but imposes a condition that hardly any sentences meet. In a sorites series of men from tall to short, there are more tall men than definitely tall men, more definitely tall men than definitely definitely tall men, and so on. Each iteration of 'definitely' reduces the number until none is left. Since the series is finite, such a point will be reached. Thus no man is definitely* tall. This conclusion does not conflict with common sense; it is not disputed that many men are definitely tall. In loose talk we may use repetition for mere emphasis, losing sight of the fact that 'definitely definitely tall' is stronger than 'definitely tall' in content as well as tone, but it remains a fact.

On the former reply, the definitely* tall men fade into the not definitely* tall men. On the latter, there are no definitely* tall men. In either case there is no sharp line between definitely* tall men and not definitely* tall men. More generally, the two replies agree that there is no sharp line between two phenomena, the perfectly straightforward application of a term and its less than perfectly straightforward application. This conclusion is anyway forced by the view that vagueness does not usually involve hidden boundaries. For a sharp boundary between the perfectly straightforward applications of a vague term and its less than perfectly straightforward applications would usually be hidden, as one can ascertain by trying to find it.

The difficulty comes out in Kit Fine's suggestion 'Anything that smacks of being a borderline case is treated as a clear borderline case'.[34] Suppose that the proposal succeeds in drawing a sharp line around the borderline cases. Then there are non-borderline cases very close to the line; but they will smack of being borderline cases, being reminiscent in appearance of cases just the other side of the line, and so count as clear borderline cases after all, which is a contradiction. Thus the proposal does not succeed. Fine's ruling extends the area of borderline cases, which extends the area of cases that smack of being borderline cases, which extends the area of borderline cases, which extends . . . ; the process has no stable limit short of including all cases. There is also the simple point that a case to which a term perfectly straightforwardly applies might smack of being a borderline case; even the wholly innocent can incur suspicion.

The term 'perfectly straightforward application of a term' is itself vague. Not even iterating the supervaluationist construction into the transfinite will give it a precise sense. There is no good reason to treat its vagueness differently from that of other terms. If their vagueness involves indeterminacy, then so does its. Supervaluationism cannot eliminate higher-order vagueness. It must conduct its business in a vague meta-language.

5.7 TRUTH AND SUPERTRUTH

In acknowledging higher-order vagueness, the supervaluationist acknowledges the vagueness of the concept of supertruth. Supertruth is truth on all admissible valuations, and the concept of admissibility is vague. This point indirectly calls into question the supervaluationist equation of truth with supertruth.[35]

Truth is standardly assumed to have the disquotational property to which Tarski drew attention. 'Cascais is in Portugal' is true if and only if Cascais is in Portugal. More generally: '*A*' is true if and only if *A*. Here '*A*' may be replaced by a sentence of the object-language under study; a truth predicate for the object-language has been added to that language to extend it to a meta-language for it. The 'if and only if' is just the material biconditional. How much more there is to the concept of truth than the disquotational property is far from clear, but in most contexts truth is assumed to be at least disquotational, whatever else it is or is not.

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, that 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. Indeed, this is just to deny the meta-linguistic equivalent of the claim that 'definitely' is a redundant operator, which the supervaluationist has already denied.

The supervaluationist did allow the statement that definitely *A* to entail and be entailed by the statement that *A*.[36] In the same way, the supervaluationist may allow the statement that '*A*' is supertrue to entail and be entailed by the statement that *A*. Were 'if and only if' to be used for mutual entailment, the disquotational schema would have a reading acceptable to the supervaluationist. It is not Tarski's reading, on which 'if and only if' is the material biconditional. More important, the mutual entailment reading fails to capture the disquotational idea. If the truth predicate really does have the effect of stripping off quotation marks, then the material biconditional that '*A*' is true if and only if *A* strips down to the tautology that *A* if and only if *A*. The supervaluationist denies that supertruth behaves like that; the availability of the mutual entailment reading is an irrelevance.

A disquotational form of truth can be introduced within the supervaluationist framework. Add quotation marks and a predicate 'true$_T$' of object-language sentences to the object-language, and let '"*A*" is true$_T$' be true on an interpretation if and only if *A* is true on that interpretation. The supervaluationist allows that either '*A*' is true$_T$ or 'Not *A*' is true$_T$, for this is to allow no more than that either *A* or not *A*. In Fine's phrase, the vagueness of 'trues$_T$' waxes and wanes with the vagueness of the given sentence. He suggests that 'true$_T$' is conceptually prior to 'supertrue', for 'supertrue' is definable in terms of 'true$_T$' and 'definitely' – a sentence is supertrue just in case it is definitely true$_T$ – and no reverse definition is possible.[37]

Truth$_T$ is disquotational; supertruth is not. In order of definition, truth$_T$ is primary; supertruth is secondary. Why then does the supervaluationist identify ordinary truth with supertruth rather than with truth$_T$? The idea seems to be that truth$_T$ is not a determinate condition, and therefore has no place in an objective semantics. Truth$_T$ is disqualified because not every sentence is either definitely true$_T$ or definitely not true$_T$. But this disqualification rests on the hopeless demand for a precise meta-language. Once higher-order vagueness is recognized, it disqualifies supertruth just as first-order vagueness disqualifies truth$_T$; not every sentence is either definitely supertrue or definitely not supertrue.[38] There is no more reason to equate ordinary truth with supertruth, definite truth$_T$, than with definite definite truth$_T$. There is more reason to identify it with truth$_T$. Truth$_T$ is vague, but so is any notion of truth we can grasp. Perhaps the ordinary concept of truth *should* match the vagueness of the sentences to which it is applied.

Once the supposed advantages of supertruth are seen to be illusory, it becomes overwhelmingly plausible to equate ordinary truth with the property that meets Tarski's disquotational condition, truth$_T$.[39] Even in a borderline case it is allowed that a vague sentence or its negation is true$_T$; thus it is either true or false in the ordinary sense. Vague sentences are not counter-examples to bivalence. Moreover, if truth is truth$_T$ rather than supertruth, then validity is a matter of preserving truth$_T$. This immediately restores the classically valid patterns of reasoning that must be abandoned if validity is a matter of preserving supertruth: contraposition, conditional proof, argument by cases, *reductio ad absurdum.*[40]The logic of 'definitely' ceases to be distinctive, becoming isomorphic in both theorems and rules of inference to a weak modal logic. What then remains of supervaluationism?

There remains the 'definitely' operator, with its semantics of admissible interpretations. However, this apparatus has lost its privileged connection with the concept of truth. Of any admissible valuation, we can ask whether it assigns truth to all and only the true sentences of the language and falsity to all and only the false ones. At most one valuation has that property. But then any other valuation will assign truth-values incorrectly, so how can it be admissible? It might be replied that no interpretation is definitely the one with the desirable property. Once definiteness has been separated from truth, that reply is without force. If an interpretation does have the desirable property, why should it matter if it does not definitely have it? Indeed, the reply is in danger of losing its sense as well as its force. If we cannot grasp the concept of definiteness by means of the concept of truth, can we grasp it at all? No illuminating analysis of 'definitely' is in prospect. Even if we grasp the concept as primitive, why suppose it to be philosophically significant?

One can make sense of the supervaluationist apparatus if one assumes that an interpretation *s* admits an interpretation *t* just in case if *s* were correct then speakers of the language could not know *t* to be incorrect. On this view, 'definitely' means something like 'knowably'. Just one interpretation is correct, but speakers of the language cannot know all others to be incorrect. Vagueness is an epistemic phenomenon. But that is not the supervaluationist view. Of supervaluationism, nothing remains articulate.