6 2

Logique & Analyse 125-126 (1989), 3-14

ARE ALL TAUTOLOGIES TRUE?

Philip HUGLY and Charles SAYWARD

More specifically, we are asking: are all tautologies true in a language with truth-value gaps? Our answer is that they are not. No tautology is false, but not all tautologies are true. Likewise, not all contradictions are false in a language with truth-value gaps, though none is ever true.

1. *Truth-value gaps vs. multivalues*

That English has truth-value gaps is trivial since English contains interrogative sentences, for example. To avoid such trivialization let us confine our attention to first-order languages (=languages or fragments thereof all of whose sentences are declarative and all of which can be put into first-order notation). That first-order English (the fragment of English satisfying the above condition) has truth-value gaps is not a trivial claim. It comes to this: First-order English contains sentences such that neither they nor their negations are true. In agreement with a number of other philosophers, we take it that first-order English has such gaps.

The claim that a language has truth-value gaps is often confused with the claim that is multivalued. But it is obvious on reflection that there is a distinction to be made here: the circumstance of there being more than two truth-values does not coincide with that of there being sentences altogether lacking truth-values. For instance, the claim that 'France is hexagonal' lacks a truth-value is not the same as the claim that it is only approximately true. In fact, the claims are incompatible, for if 'France is hexagonal' is approximately true then it does have a truth-value — approximate truth — in which case it does not lack a truth-value.

How else might a language be multivalued? Well, if there really are such things as half-truths, for example, then a language whose sentences fall into the three categories of full truths, half-truths, and falsehoods, is a 3-valued language and thus multivalued. (And such a language might also contain sentences which are neither fully nor partially true, nor false, nor in any other way truth-valued.) Or it might be that there are infinitely many truth-values ranging from 1 (complete truth) to 0 (complete falsity), as has been proposed by some philosophers as a way of dealing with sorites paradoxes.(1) These would both be cases of multivalues consisting of different *degrees* of truth and falsity. Another way a language might be multivalued would be for it to have various *kinds* of truth and falsity. It is sometimes suggested that 'Bachelors are unmarried' has a "positive" truth-value of a kind distinct from the "positive" truth-value of 'Snow is white'. Other candidates for different kinds of truth-values are the truth-values of moral sentences (moral truth and moral falsity) or of mathematical sentences (mathematical truth and mathematical falsity).

A language with truth-value *gaps*, on the other hand, is one which has sentences which have *no* truth-values at all; such sentences would not only not be true or false to any degree, they also would not have any kind of truth or falsity.

The question we wish to address is that of how to understand the usual sentential connectives treated of in classical logic when the assumption of bivalence is waived for two-valued languages. That is, we wish to determine the right account of the connectives for the case of languages with just two truth-values some sentences of which lack either. On the account we shall defend, not all tautologies are true.

2. *The Kleene tables*

S.C. Kleene proposed evaluation tables for the usual sentential connectives intended to capture their use in mathematical contexts.(2) His table for negation is

A	¬A
t	f
u	u
f	t

and his tables for disjunction and conjunction are

(1) J. Goguen, "The Logic of Inexact Concepts" *Synthese* 19 (1964), 325-73.

(2) S. C. Kleene, *Introduction to Metamathematics*, North Holland (1952), pp. 332-340.

$\vee$	t	u	f
t	t	t	t
u	t	u	u
f	t	u	f

and

$\wedge$	t	u	f
t	t	u	f
u	u	u	f
f	f	f	f

He construes the conditional $A \rightarrow B$ as $\neg A \vee B$ and the biconditional $A \leftrightarrow B$ as $(A \rightarrow B) \wedge (B \rightarrow A)$, so that the tables for these connectives are

$\rightarrow$	t	u	f
t	t	u	f
u	t	u	u
f	t	t	t

and

$\leftrightarrow$	t	u	f
t	t	u	f
u	u	u	u
f	f	u	t

How are these tables to be understood? For what do the marks 't', 'f', and 'u' stand? That mathematical sentences may admit of degrees of truth and falsity lacks plausibility and has not, to our knowledge, as yet been argued for. So it is unlikely that 'u' should be thought of as standing for some intermediate truth-value. That mathematical sentences admit of various kinds of truth or falsity seems equally implausible. So it is unlikely that, say, 'u' and 't' should be thought of as standing for two different kinds of truth.

Nonetheless, it is not implausible to think that mathematical sentences admit of more than two values. The conception which suggests this conclusion runs roughly as follows: Mathematical sentences are either *provable*, *disprovable*, or *undecidable*, with the provable sentences being all true, the disprovable sentences being all false and the undecidable sentences being those whose truth-value is indeterminable. In the spirit of this conception, Susan Haack has set out the following suggestion as to how Kleene intended his tables to be understood:

> ...Kleene so interprets 'u' ('undecidable') that sentences which take u are nevertheless either true or false though it is not possible to tell which. Thus we have
> t = true (certainly)
> u = true or false (but undecidable which)
> f = false (certainly)
> so that once again 'true' and 'false' are jointly exhaustive (3).

On this view truth, falsity and undecidability really are three *different* values even though no mathematical sentence has the third of these values without also having one of the first two of these values. But this does not cohere with Kleene's tables, since A → A is provable, and thus decidable, even if A is neither provable nor disprovable; but by Kleene's tables, if A takes u then so does A → A. An alternative view, which we think best interprets Kleene's tables, is one on which mathematics is two-valued but not bivalent in the sense that the only values attaching to mathematical sentences are two values of truth and falsity (=mathematics is two-valued) but that some mathematical sentences lack both of these values (=mathematics is not bivalent). We thus interpret u as marking a truth-value gap. A good way of notationally marking this interpretation is to use in place of 'u' a blank space or, perhaps, a dash signifying a gap. For example, writing

→	t	-	f
t	t	-	f
-	t	-	-
f	t	t	t

(3) Susan Haack, *Deviant Logic*, Cambridge University Press (1974), p. 61.

would help to make it clear that we are not thinking in terms of how three values distribute over the conditional, but of how the conditional turns out when one or more of its component sentences lacks a truth-value.

How are tautologies and contradictions to be understood if the connectives are interpreted by the Kleene tables? Clearly, no schema gets only t under its main connective, and none gets only f under its main connective. Instead, there are some schemata which never get f under their main connectives and some which never get t under their main connectives. The former are exactly the tautologies of classical logic and the latter are exactly the truth-functional inconsistencies of classical logic. Consequently, a tautologous schema is simply one which is never assigned f, so that all tautologies are true if truth-valued, though some are not true since not truth-valued. And the converse holds for contradictions. Thus, some tautologies and some contradictions lack a truth-value. For example, $A \rightarrow A$ and its negation lack a truth-value if A does. Still, $A \rightarrow A$ is a tautology and its negation is a contradiction.

Why do we find it plausible to interpret Kleene's 'u' as marking the *lack* of a truth-value? It is because his tables are uniquely determined by that assumption together with three highly plausible principles for two-valued languages with value gaps.

Let L be a two-valued language. The principles are these:

P1 The classical rules of inference governing the connectives hold for L.
P2 These connectives are truth-functional in L.
P3 These connectives never form truth-valued sentences from sentences all of which lack truth-values.

(If, in addition to being two-valued, L has gaps, then truth-functionality comes to this: Let X be a sentence $C(A_1,\ldots,A_n)$ constructed from component sentences $A_1,\ldots,A_n$ and n-ary connective C in some usual way; let Y be $C(B_1,\ldots,B_i)$; and let V be an evaluation function mapping a subset of the sentences of L onto the set of truth-values. Then C is truth-functional if, and only if, whenever $V(A_i) = V(B_i)$ or both $V(A_i)$ and $V(B_i)$ are undefined $(1 \leq i \leq n)$, $V(X) = V(Y)$ or both $V(X)$ and $V(Y)$ are undefined.)

P1 requires that double negation

$$A \,/\, \neg\neg A \qquad \neg\neg A \,/\, A$$

be truth preserving. This and P2 requires that the negation table yields a truth from the negation of a falsity and a falsity from the negation of a truth.(4) By P3 that table must also withhold truth-values from the negation of what lacks a truth-value. P1 - P3 thus fix the Kleene negation table for a two-valued language with truth-value gaps.

The case for disjunction is similar. P3 tells us that if A and B both lack truth-values, then the same must hold for their disjunction, A ∨ B. P1 and P2 forces the remaining entries since the following inference rules for disjunction and negation must be truth preserving:

$$A / A \vee B \quad B / A \vee B \quad \neg A, A \vee B / B \quad \neg B, A \vee B / A$$

$$\neg(A \vee B) / \neg A \quad \neg(A \vee B) / \neg B$$

The entries for the other connectives are determined by their familiar reduction to disjunction and negation.

A comment may be in order here. It might seem natural in connection with P1 and P2 to say that they insure that the tables to which they lead must validate classical rules of inference in the sense of making them *truth-preserving*. But those rules, applied in a two-valued language with gaps, make it possible to derive any tautology from any sentence and thus to derive non-truth-valued sentences from true sentences. (Cf. the last paragraph of this section.)

We also note the objection that it is sometimes said that a tautology like 'Socrates was wise if Socrates was wise' is true independently of how the world is, whereas a sentence like 'Plato was wise if Socrates was wise' is true in virtue of how the world is. This leads to the idea that certain sentences are true in virtue of their structure alone. But that seems to us to be an error. In fact the two sentences just instanced both have their truth in virtue of how the world is. Both are conditionals with true consequents and which thus, like any conditionals with true consequents, owe their truth to that.

(4) An anonymous referee of this journal has called our attention to the following possibility: Let $V(A)=1$ just in case A is a theorem of classical propositional logic; otherwise $V(A)=0$. Then double negation is "truth" preserving in that $V(A)=1$ just in case $V(\neg\neg A)=1$. And, while $V(A)=1$ entails that $V(\neg\neg A)=0$, the converse does not hold. (Cf. Nuel D. Belnap and Gerald J. Massey, "Semantic Holism", *Studia Logica*, XLIX, 1, 1990, pp. 67-82; Gerald J. Massey, "Semantic Holism is Seriously False", *Studia Logica*, XLIX, 1, 1990, pp. 83-86.). Note that the connectives are not truth-functional on this construal. This is where P2 comes in.

And that consequent — the sentence 'Socrates was wise' — is true in virtue of the *fact* that Socrates was wise. So, both conditionals are true in virtue of how the world is. The conditionals 'Socrates was unjust if Plato was unjust' and 'Socrates was unjust if Socrates was unjust' are also true in virtue of how the world is. For these conditionals owe their truth to the falsity of their antecedents — the sentence 'Socrates was unjust' — and this sentence is false in virtue of the circumstance that Socrates was just. Certain sentences have a connective structure which insures truth no matter how their component sentences turn out true or false. But no sentence has a connective structure which *itself* makes for truth. If we were to form a conditional from two occurrences of a sentence which entirely lacks any truth-value, then its mere syntactic structure would not avail to yield a truth-value. Interposing a connective between a pair of occurrences of a sentence lacking a truth-value cannot produce truth *or* falsity.

As noted earlier, given these tables there will be certain schemata no instances of which are false. These are exactly the tautological schemata of classical logic and their instances are exactly the tautologies of classical logic. Parallel points hold for contradictions.

We here view classical logic as the logic of a *special case*, the case of two-valued languages *lacking gaps*. The *general case* is that of two-valued languages, with or without gaps. Correspondingly, the *general* conception of a tautology is that of a sentence whose connective structure precludes falsity. For the *special* case of a two-valued language *lacking* gaps, the structures which preclude falsity thereby insures truth. But what insures truth then is not structure alone. And, in similar fashion, for the special case of two-valued languages lacking gaps validity of inference insures that rules of inference are truth-preserving. But the *general* conception of a valid inference is one which is, as it were, falsity-avoiding; it is an inference of a type which cannot yield falsity from truth alone.

3. *Supervaluations*

Truth-value gaps pose interesting questions about what is and is not central to logic. On the view just set out, truth-value gaps show us that what is central to our *general* conception of a logically valid sentence is not that it is one whose structure makes for truth, but that it is one whose structure precludes falsity.

On another view of what is and is not central to logic — the view set forth by Bas C. van Fraassen in his paper on supervaluations(5) — truth-value gaps fail to show that this is so. If van Fraassen is right, a logically valid sentence is one whose structure insures truth *even if* that structure embeds only truth-valueless sentences. That is, on his view connective structure alone is in certain instances sufficient to generate truth. It is, he thinks, central to logic that certain sentences are logical *truths* (in the sense of being true no matter what) and that certain other sentences are logical *falsehoods* (in the sense of being false no matter what).

Applied to the case of sentential logic, this conception implies that it is central to logic that all tautologies are true (no matter what) and that all their negations (here to be called "contradictions") are false (no matter what). If so, then the tautologies and contradictions of languages with truth-value gaps are nonetheless respectively true and false, even when their components are one and all without truth-values.

In particular, supposing with van Fraassen that 'Pegasus has a white hind leg' and its negation are both neither true nor false, the tautologous sentence 'Pegasus has a white hind leg or it is not the case that Pegasus has a white hind leg' is yet *true*. But is it?

Van Fraassen gives only one argument that this is so. The argument is couched in terms of his discussion of supervaluations. He there argues that there is a method of assigning T, F or nothing at all to sentences which assigns T to this disjunction but assigns neither T nor F to its disjuncts. This is the method of supervaluations. The underlying idea or assumption is that the assigned values T and F respectively correspond to truth and falsity — that being assigned T (F) by an appropriate supervaluation formally represents the circumstance of being *true* (or *false*).

Van Fraassen works from the idea that an *atomic* sentence is true if its names all succeed in naming and its predicate is true of the things named in the order in which they are named; it is false if all the names name but the predicate is false of these things in the order named; and it is neither true nor false if one or more of the names fails to actually name. Our own remarks will fully grant this idea.

The circumstance just reviewed is formally represented by a language L and a model $\langle f,D \rangle$ of L where D is a non-empty domain set and f is a function fully defined in the usual way for the predicates of L and partially

(5) Bas C. van Fraassen, "Singular Terms, Truth-Value Gaps and Free Logic", *Journal of Philosophy* 63 (1966), 481-495.

defined in the usual way for the names of L. An *atomic* sentence Fa...n of L is *true (false)* in the model ⟨f,D⟩ iff (i) f is defined for the names a...n and (ii) ⟨f(a),...,f(n)⟩ is a member of f(F) (is not a member of f(F)). For negations and disjunctions van Fraassen invokes a further function S as follows: If A is the negation of a sentence B, then A is true (false) in ⟨f,D⟩ iff S(B)=T (or S(B)=F), and if A is the disjunction of sentences B and C, then A is true (false) in ⟨f,D⟩ iff either S(B)=T or S(C)=T (or S(B)= S(C)=F).

The function S is defined in terms of classical valuations over models. A classical valuation over a model ⟨f,D⟩ is a function which, for each atomic sentence A of L, assigns T to A if A is true in ⟨f,D⟩ and assigns to F to A if A is false in ⟨f,D⟩ and assigns one of T or F to A if A is neither true nor false in ⟨f,D⟩, and assigns T and F to the remaining sentences in accord with the usual rules. A supervaluation over a model ⟨f,D⟩ assigns T or F to an atomic sentence as that sentence is true or false in ⟨f,D⟩ and otherwise assigns T to a sentence iff that sentence is assigned T by every classical valuation over ⟨f,D⟩ and assigns F to a sentence iff that sentence is assigned F by every classical valuation over ⟨f,D⟩.

To see what this looks like, we consider a particular example: the sentence

$$(Fa \vee Fb)$$

and model ⟨f,D⟩ where f is defined for neither 'a' nor 'b'. Then both disjunct sentences lack truth-values in ⟨f,D⟩. The classical valuations on ⟨f,D⟩ will thus assign T and F to these disjuncts in all possible ways. So there is a classical valuation assigning T to the disjunction and also a classical valuation assigning F to the disjunction. Thus, the supervaluation S assigns nothing to that disjunction. On the other hand, for the disjunction

$$(Fa \vee \neg Fa)$$

every classical valuation on ⟨f,D⟩ assigns T to one disjunct and F to the other. Thus, the supervaluation S assigns T to the disjunction. And so will all supervaluations over all models of L. So this tautology — despite having only components which are neither T nor F — is not only assigned T by the supervaluation over ⟨f,D⟩ but by all other supervaluations as well.

It is clear that the function S over ⟨f,D⟩ assigns T (F) to an atomic sentence just in case that sentence is true (false) in ⟨f,D⟩. Thus, the circumstance that S assigns T (F) to an *atomic* sentence formally represents the circumstance of that atomic sentence being *true* (*false*). Does this hold more generally? Is it right to regard the assignments of T and F as formally representing truth and falsity for non-atomic sentences?

Since supervaluational assignments proceed via classical assignments, let us look at those first. Let the language be L where 'Fa' is a sentence of L and 'a' is a non-denoting name of L. In that case, the sentence 'Fa' is neither true nor false. The model ⟨f,D⟩ together with the definition of truth in ⟨f,D⟩ for atomic sentences of L represents this situation in formal terms. Among the classical valuations definable on ⟨f,D⟩ is one which assigns T to 'Fa'. What does *this* represent? Certainly not that 'Fa' *is* true, for it is not — since 'a' fails to denote. Rather, what is represented is the circumstance that it is possible for 'F' and 'a' to have denotations on which 'Fa' would be true. What the classical valuation represents is not truth (falsity) for 'Fa', but the *possibility* of truth (falsity) for 'Fa'. The sentence 'Pegasus has a white hind leg' is neither true nor false, though it *could* be made true by assigning the name 'Pegasus' to something with a white hind leg or could be made false by assigning that name to something with, say, a black hind leg. In particular, that some classical valuation assigns T to 'Pegasus has a white hind leg' does not formally represent the circumstance that this sentence is true (since it *isn't* true) but only the circumstance that its syntactic form is no bar to its being true.

Consider now the sentence '(Fa ∨ ¬Fa)' when 'Fa' is without a truth-value due to the lack of denotation for 'a'. What then of '¬Fa'? Will it be true, be false, or be neither? Presumably neither. Certainly, it will be assigned neither T nor F by the supervaluation defined on ⟨f,D⟩. What then is the truth-value status of the disjunction? The supervaluation S defined on ⟨f,D⟩ assigns T to this disjunction. But that is just to say that all classical valuations over ⟨f,D⟩ assign it T which in turn says only that every possible way for 'Fa' and '¬Fa' to turn out truth-valued makes the disjunction true. And *that* is not to say that the disjunction *is* true.

That '(Fa ∨ ¬Fa)' is assigned T by S indicates only that *if* 'Fa' is made truth-valued by determining denotations for its parts, then the disjunction will be made true by those determinations. It does not indicate that the sentence actually *is* true.

The point at hand is very simple. It comes just to this, that '... is true' and '... can turn out only true if all its names are assigned denotations' do *not* come to the same. Being true is one thing and being true *if* all denotation "gaps" are filled is another. Indeed, what seems plainly to be the case is that since 'a' lacks denotation, the disjunction '(Fa ∨ ¬Fa) is *not* true. That this is a sentence of a form which yields only truths from truth-valued components in no way argues for the opposite.

van Fraassen gives just one argument that the T-assignment of the supervaluation represents truth in the model. It runs as follows:

> ...the classical valuations go beyond the model to which they belong just with respect to those terms which have no referent in the model. But there they go beyond the model *in all possible ways*...So what the classical valuations have in common we can take as correctly reflecting truth and falsity in the model.([6])

So the idea is that since all classical valuations of a tautology assign it 'T', tautologies are true in a model even if their components are neither true nor false in the model. But why suppose that this is so?

Suppose that several groups of people speak in the syntax of English and so use the terms thereby made available that they do not differ in denotation except on the name 'Pegasus'. In group A's use of the word 'Pegasus' it fails to denote. In group B's use of that name it denotes a particular white legged animal. In group C's use of that name it denotes a particular black legged animal. Now, no one supposes that since 'Pegasus has a white hind leg' is true in B-English it also is true in A-English, or that since it is false in C-English it is false in A-English. van Fraassen would entirely agree with this. But, he holds, the circumstance that 'Pegasus has a white hind leg if Pegasus has a white hind leg' is true in B-English and is true in C-English and is true in any English in which 'Pegasus' has a denotation "reflects" that this sentence also is true in A-English. But this is merely asserted.

What clearly is the case (and what comes out sharply in the natural account of the Kleene tables given above) is that the connective structure of 'Pegasus has a white hind leg if Pegasus has a white hind leg' precludes falsity and insures truth in any possible variants of English in which its component sentence is truth-valued. But this provides no support for the claim that the conditional is true in the English we actually speak, the one in which 'Pegasus has a white hind leg' is neither true nor false. That a sentence can only be true if its components are truth-valued simply is not

([6]) *Ibid.* p. 483.

the same as its being true when its components lack truth-values. van Fraassen's argument comes down to an inference from the first point to the second. It is an inference we see no reason to make.

University of Nebraska-Lincoln