Logic for Equivocators

DAVID LEWIS

PRINCETON UNIVERSITY

I

At first the champions of relevance in logic taught that relevance and preservation of truth were two separate merits, equally required of any relation that claims the name of implication. Thus Anderson and Belnap, in [1]: 19-20, cheerfully agree that certain classical implications necessarily preserve truth, but fault them for committing fallacies of relevance. (And there is no hint that necessary truth-preservation might not be truth-preservation enough.) Lately, however, relevance has been praised not—or not only—as a separate merit, but rather as something needed to ensure preservation of truth. The trouble with fallacies of relevance, it turns out, is that they can take us from truth to error.

Classical implication does preserve truth, to be sure, so long as sentences divide neatly into those that are true and those with true negations. But when the going gets tough, and we encounter true sentences whose negations also are true, then the relevant logician gets going. Then his relevant implication preserves truth and some classical implication doesn't. Say that A is true and its negation $\sim A$ is true as well. Then $A \ \& \sim A$ is a conjunction of truths, and hence true. Let B be false, and not true. Then the classically valid, but irrelevant, implication *ex falso quodlibet*

$$\frac{A \ \& \sim A}{\therefore B} \qquad \text{or} \qquad \frac{A \quad \sim A}{\therefore B}$$

fails to preserve truth. Likewise $A \vee B$ is true, being a disjunction with a true disjunct. So disjunctive syllogism, the relevantist's bugbear,

$$\frac{\begin{array}{c} A \vee B \\ \sim A \end{array}}{\therefore B}$$

also fails to preserve truth in this case. Whether or not we love relevance for its own sake, surely we all agree that any decent implication must preserve truth. So should we not all join the relevantists in rejecting *ex falso quodlibet* and disjunctive syllogism as fallacious? So say Belnap [2], Dunn [3] and [4], Priest [10], Routley [12] and [13], and Routley and Routley [14]. See also Makinson [7], Chapter 2, who explains this motiviation for relevance without himself endorsing it.

The proposed vindication of relevance can be spelled out with all due formality. But first we must settle a verbal question. If there are truths with true negations, what is it to be false? Shall we call a sentence false if and only if it has a true negation? Or if and only if it isn't true? (In the terminology of Meyer [8]: 3-6, shall we use "American" or "Australian" valuations? In the terminology of Routley [13]: 329, note 11, shall we speak of "systemic" or "metalogical" falsity? Or rather—since we can speak of both and needn't choose—which shall get the short name "falsity"?) Like the majority of our authors, I shall speak in the first way. A truth with a true negation is false as well as true. Its negation is true (as well as false) because negation reverses truth values. We shall have a three-valued semantics; or perhaps four-valued, if we allow truth-value gaps. Two of the values are regarded as two sorts of truth: a sentence is true if and only if it is either true only or both true and false. Likewise we have two sorts of falsity: some false sentences are false only, some are both true and false. Perhaps also some sentences are neither true nor false.

Let our language be built up from atomic sentences using just the standard connectives $\sim$, &, and $\vee$, governed by these orthodox rules of truth and falsity.

(For $\sim$) $\sim A$ is true iff A is false, false iff A is true.
(For &) A & B is true iff both A and B are true,
false iff either A or B is false.
(For $\vee$) $A \vee B$ is true.
iff either A or B is true,
false iff both A and B are false.

No classical logician could call these rules mistaken, though some might call them longwinded. But they must be longwinded if they are to apply when sentences are both true and false, or when they are neither. We may check that according to these rules, as we expected, *ex falso quodlibet* and disjunctive syllogism fail to preserve truth when A is both true and false and B is false only. Their premises are, *inter alia*, true, their conclusions are not. The story continues in three versions.

Version E (Dunn [3]; Belnap [2]; and Makinson [7]: 30-8 on "De Morgan implication"). We allow truth-value gaps, and thus we have a four-valued semantics. A premise set implies a conclusion if and only if any valuation that makes each premise true also makes the conclusion true. The implications so validated turn out to be just those given by the first-degree fragment of the Anderson-Belnap logic E of entailment. Along with *ex falso quodlibet* and disjunctive syllogism, the alleged fallacies of relevance

$$(*)\quad \frac{A}{\therefore B \vee \sim B} \qquad\qquad (**)\quad \frac{A \;\&\; \sim A}{\therefore B \vee \sim B}$$

also are not validated. They fail in case A is both true and false and B is neither.

Version RM (Dunn [4]; Dunn [3]: 166-7, note 7; and Makinson [7]: 38-9 on "Kalman implication"). We prohibit truth-value gaps and require valuations to make every sentence true or false or both; thus we have a three-valued semantics. We complicate the definition of implication, requiring not only truth but also truth-only to be preserved: (i) any valuation that makes each premise true also makes the conclusion true, and (ii) any valuation that makes each premise true and not false also makes the conclusion true and not false. The implications so validated are those given by the first-degree fragment of the partly relevant logic R-mingle. The irrelevant (**) is validated; but *ex falso quodlibet*, disjunctive syllogism, and (*) are not. Note that (*) fails for a new reason: if A is true only and B is both true and false, then (*) preserves truth but fails to preserve truth-only.

Version LP (Priest [10]). Again we have only three values: true only, false only, both true and false. But we return to the simple definiton of implication, no longer requiring preservation of truth-only. As in Version E, a valid implication is one such that any valuation that makes each premise true also makes the conclusion true. The implications thus validated are given by the first-degree fragment of Priest's "logic of paradox" LP. Both (*) and (**) are validated, so the vindication of relevance on this version is very partial indeed; but the principle offenders, *ex falso quodlibet* and disjunctive syllogism, still fail. We pay dearly for our simplified definition of truth-preserving implication, since we lose contraposition. Although A implies $B \vee \sim B$ (irrelevantly), $\sim(B \vee \sim B)$ does not contrapositively imply $\sim A$: for a counterexample, let B be both true and false and let A be true only. Thus Priest's LP is weaker than E when it comes to contraposition, but stronger even than R-mingle when it comes to tolerating some alleged fallacies of relevance.

II

All this may help illuminate the technicalities of relevant logic; but of course it is worthless as an intuitive vindication of relevance unless somehow it makes sense that sentences can be both true and false. How can that be? Two answers are on the market. I am persuaded by neither of them.

Routley [12] and [13] and Priest [10] offer a radical answer. Maybe no special explanation is needed, but only liberation from traditional dogma. Maybe some truths just *do* have true negations. Maybe some sentences just *are* both true and false. (For Routley: "systemically" false.) Routley and Priest mention various candidates, and others come to mind as well. (i) We have our familiar budget of logical paradoxes: the liar, the Russell set, and so on. We have very persuasive arguments that the paradoxical sentences are true, and that they are false. Why not give in, stop struggling, and simply assent to both conclusions? (ii) On some subjects, the truths and falsehoods are of our own making. The Department can make it true or false by declaration that a dissertation is accepted. If by mistake or mischief both declarations were made, might both succeed? (iii) Thorny problems of physics might prove more tractable if we felt free to entertain inconsistent solutions—the particle has position p, it has momentum q, but it does not have both position p and momentum q. (iv) Likewise for thorny problems of theology. Indeed, this proposal amounts to a generalization to other realms of the plea that "God is not bound by human logic".

This radical answer leads *prima facie* to a three-valued semantics, Version RM or Version LP. To be sure, it could be combined with provision for truth-value gaps—and intuitive motivations for gaps are a dime a dozen—to yield Version E. But it does not seem to provide a single, uniform motivation for the four-valued semantics. Even if somehow there are indeterminacies in the world as well as contradictions, there is no reason to think that the two phenomena are two sides of the same coin.

The reason we should reject this proposal is simple. No truth does have, and no truth could have, a true negation. Nothing is, and nothing could be, literally both true and false. This we know for certain, and *a priori*, and without any exception for especially perplexing subject matters. The radical case for relevance should be dismissed just because the hypothesis it requires us to entertain is inconsistent.

That may seem dogmatic. And it is: I am affirming the very thesis that Routley and Priest have called into question and—contrary to the rules of debate—I decline to defend it. Further, I concede that it is indefensible against their challenge. They have called so much into question that I have no foothold on undisputed ground. So much the worse for the demand that philosophers always must be ready to defend their theses under the rules of debate.

***

There is another, more conservative, answer to the question how a sentence can be both true and false. It is suggested by Dunn [3]; it appears alongside the radical answer to Routley [12]; and some of its ingredients appear in Belnap [2]. (Belnap discusses the problem of quarantining inconsistencies in the data bank of a question-answering computer, but declines to suggest that we and the computer well might solve our similar problems in similar ways.) This answer runs as follows. Never mind "ontological" truth and falsity, that is, truth and falsity *simpliciter*. Instead, consider truth and falsity according to some corpus of information. It might be someone's system of beliefs, a data bank or almanac or encyclopedia or textbook, a theory or a system of mythology, or even a work of fiction. The information in the corpus may not all be correct, and misinformation may render the corpus inconsistent. Then we might well say that sentences about matters on which there is an inconsistency are both true according to the corpus and false according to the corpus; and we might say the same of their negations. We want a conception of truth according to a corpus such that, if the corpus is mostly correct, then truth according to it will serve as a good, if fallible, guide to truth *simpliciter*. We can reasonably ask that such a conception satisfy the following desiderata. (1) Anything that is explicitly affirmed in the corpus is true according to it. (2) Truth according to the corpus is not limited to what is explicitly there, but is to some extent closed under implication. (This may include implication with the aid of background information, but let us ignore this complication.) (3) Nevertheless, an inconsistency does not—or does not always—make everything true according to the corpus. Hence truth according to the corpus is closed not under classical implication generally, but under some sort of restricted implication capable of quarantining inconsistencies. Maybe not all inconsistencies can be quarantined successfully, but many can be. (4) A sentence is false according to the corpus if and only if its negation is true according to the corpus.

This proposal looks as if it could give us an intuitive, uniform motivation for the four-valued semantics of Version E and for the resulting case against irrelevant implication. (Presumably it would not work to motivate the three-valued versions. A corpus is even more likely to be incomplete than inconsistent, so it is inevitable that some sentences will be neither true nor false according to it. To avoid such gaps, we would have to pass from real-life corpora to their completions.)

But in fact, I do not think the proposal succeeds. I agree with much of it. If a corpus is inconsistent, I see nothing wrong with saying that a sentence and its negation both may be true according to that corpus; or that such a sentence is true according to it and also false according to it. I further agree that we have a legitimate and useful conception of truth according to a corpus that satisfies the four desiderata I listed. So far, so good. But the conception I have in mind does not work in a way that fits the four-valued semantics. Nor does it use restrictions of relevance to quarantine inconsistencies. Instead, it uses a method of fragmentation. I shall explain this informally; for the technicalities, see Jaśkowski [6], Rescher and Brandom [11], and Schotch and Jennings [15]. (However, these authors' intended applications of fragmentation differ to some extent from mine.)

I speak from experience as the repository of a mildly inconsistent corpus. I used to think that Nassau Street ran roughly east-west; that the railroad nearby ran roughly north-south; and that the two were roughly parallel. (By "roughly" I mean "to within 20°".) So each sentence in an inconsistent triple was true according to my beliefs, but not everything was true according to my beliefs. Now, what about the blatantly inconsistent conjunction of the three sentences? I say that it was not true according to my beliefs. My system of beliefs was broken into (overlapping) fragments. Different fragments came into action in different situations, and the whole system of beliefs never manifested itself all at once. The first and second sentences in the inconsistent triple belonged to—were true according to—different fragments; the third belonged to both. The inconsistent conjunction of all three did not belong to, was in no way implied by, and was not true according to, any one fragment. That is why it was not true according to my system of beliefs taken as a whole. Once the fragmentation was healed, straightway my beliefs changed: now I think that Nassau Street and the railroad both run roughly northeast-southwest.

I think the same goes for other corpora in which inconsistencies are successfully quarantined. The corpus is fragmented. Something about the way it is stored, or something about the way it is used, keeps it from appearing all at once. It appears now as one consistent corpus, now as another. The disagreements between the fragments that appear are the inconsistencies of the corpus taken as a whole. We avoid trouble with such inconsistencies (and similar trouble with errors that do not destroy consistency) by not reasoning from mixtures of fragments. Something is true according to the corpus if and only if it is true according to some one fragment thereof. So we have no guarantee that implication preserves truth according to the corpus, unless all the premises come from a single fragment. What follows from two or more premises drawn from disagreeing fragments may be true according to no fragment, hence not true according to the corpus.

The details of the implication do not matter. Two-premise "fallacies of relevance" such as disjunctive syllogism, or the version of *ex falso quodlibet* in which A and $\sim A$ are separate premises, may indeed fail to preserve truth according to an inconsistent corpus. But so may many-premise implications of impeccable relevance, such as the implication from conjuncts to their conjunction (as in our example). Only one-premise implications can be trusted not to mix fragments. Irrelevant one-premise implications—such "fallacies" as (*), (**), and the one-premise version of *ex falso quodlibet*—can no more mix fragments than any other one-premise implications can. In short, to the extent that inconsistencies are quarantined by fragmentation, restrictions of relevance have nothing to do with it.

If we take truth values according to an inconsistent corpus, it may well happen that A and $\sim A$ are both true (as well as false) and B is false only. But that is not enough to give us a counterexample against *ex falso quodlibet* in its one-premise version. Such a counterexample requires truth according to the corpus of the inconsistent conjunction A & $\sim A$, not just of its conjuncts. But the conjunction need not be true according to the corpus. The corpus may be fragmented, and the conjuncts may be true according to different fragments. As we have already seen, conjunction need not preserve truth according to the corpus.

The four-valued semantics of Version E gave us orthodox rules of truth and falsity for the connectives. (Likewise for the three valued versions.) These rules are unexceptionable for truth and falsity *simpliciter*. But if we try to understand the four-valued semantics in terms of truth and falsity according to a corpus, and if the corpus is fragmented, then the rule for & may fail. It is such a failure that wrecks our counterexample against one-premise *ex falso quodlibet*. The rule for $\vee$ also may fail, wrecking our previous counterexamples against (*) and (**); I spare you the details. Therefore this strategy does not lead to a successful interpretation of the four-valued semantics, even though it does provide a way in which sentences can be regarded as both true and false.

I am inclined to think that when we are forced to tolerate inconsistencies in our beliefs, theories, stories, etc., we quarantine the inconsistencies entirely by fragmentation and not at all by restrictions of relevance. In other words, truth according to any single fragment is closed under unrestricted classical implication. This view would take some strenuous defending, but its defense is irrelevant to my present purpose. For if the quarantine works only partly by fragmentation, that is enough to make my point. That is enough to make the rules for & and $\vee$ fail, so that the four-valued semantics of Version E will not apply. (Nor will the three-valued semantics.) Further, even if a mixed quarantine disallows the alleged fallacies of relevance, we still get no vindication of the relevant logic E. It cannot be trusted to preserve truth according to a fragmented corpus, nor can any logic that ever lets us mix fragments in many-premise implications.

A concession. I need not quarrel with anyone who wishes to put forward a fifth desideratum for a conception of truth according to a corpus: (5) the orthodox rules for & and ∨ must apply without exception. I claim no monopoly on behalf of the conception I have been discussing. I gladly agree that there are legitimate and intuitive conceptions that do satisfy desideratum (5), along with some of the original four. One such conception is obtained by closing under unrestricted classical implication; that satisfies all the desiderata except (3). Another is obtained by taking the intersection rather than the union of the closures of the fragments, so that a sentence is true according to the corpus if and only if it is true according to every fragment; that satisfies all the desiderata except (1). What I doubt is that there is any useful and intuitive conception that satisfies all five desiderata. When asked to respect (1)-(4), I come up with a conception that violates (5). But a natural conception that satisfies all five is what the relevantist needs, if he is to vindicate relevance in the way we have been considering.

III

I conclude that if we seek a vindication of relevance, we must find a third way in which sentences can be regarded as both true and false. I suggest we look to ambiguity. (Similar suggestions appear in Dunn [5]: 167-9 and in Pinter [9]. But I am not sure that these authors mean what I do by ambiguity. Dunn illustrates it with what seems more like a case of unspecificity. Pinter defines an ambiguous proposition as one that is both true and false; if that applies to sentences at all, it makes "Fred went to the bank" unambiguous if he went nowhere, or if he went to the riverside branch of the First National.)

Strictly speaking, an ambiguous sentence is not true and not false, still less is it both. Its various disambiguations are true or false *simpliciter*, however. So we can say that the ambiguous sentence is true or is false on one or another of its disambiguations. The closest it can come to being simply true is to be true on some disambiguation (henceforth abbreviated to "osd"); the closest it can come to being simply false is to be false-osd. Barring independently motivated truth-value gaps for the disambiguations, we have just three possibilities. A sentence can be true-osd only, false-osd only, or both true-osd and false-osd. That is, it can be true on all its disambiguations, false on all, or true on some and false on others. Here we have a possible intuitive interpretation of the three-valued semantics of Versions RM and LP.

It would be hard to get the four-valued semantics of Version E. We would be trying for the fourth value: neither true-osd nor false-osd.

That makes sense if we have truth-value gaps; a sentence might be gappy on all its disambiguations. (Presumably any sentence has at least one disambiguation.) But if we have gaps, how do we end up with four values rather than seven? I can see no reason to exclude these three extra values:

gappy-osd, true-osd, not false-osd;
gappy-osd, not true-osd, false-osd;
gappy-osd, true-osd, false-osd.

Accordingly, I shall consider the three-valued versions only.

We must check that the orthodox rules for the connectives still hold when truth and falsity are understood as truth-osd and falsity-osd. They do, provided that we admit *mixed disambiguations*: that is, disambiguations of a compound sentence in which different occurrences of the same ambiguous constituent are differently disambiguated. We should admit them; they are commonplace in ordinary language. Consider the most likely disambiguation of "Scrooge walked along the bank on his way to the bank"—he fancied a riverside stroll before getting to work with the money.

It makes no sense to say that an implication involving ambiguous sentences preserves truth *simpliciter*. But it may preserve truth-osd. Also it may preserve truth-osd-only, in other words truth on all disambiguations. The implications that preserve truth-osd are those given by the first-degree fragment of Priest's LP. Those that preserve both truth-osd and truth-osd only are given by the first-degree fragment of R-mingle. So we have two logics for ambiguous sentences—and lo, they are partly relevant.

Logic for ambiguity—who needs it? I reply: pessimists.

We teach logic students to beware of fallacies of equivocation. It would not do, for instance, to accept the premise $A \vee B$ because it is true on one disambiguation of A, accept the premise $\approx A$ because it is true on another disambiguation of A, and then draw the conclusion B. After all, B might be unambiguously false. The recommended remedy is to make sure that everything is fully disambiguated before one applies the methods of logic.

The pessimist might well complain that this remedy is a counsel of perfection, unattainable in practice. He might say: ambiguity does not stop with a few scattered pairs of unrelated homonyms. It includes all sorts of semantic indeterminacy, open texture, vagueness, and whatnot, and these pervade all of our language. Ambiguity is everywhere. There is no unambiguous language for us to use in disambiguating the ambiguous language. So never, or hardly ever, do we disambiguate anything fully. So we cannot escape fallacies of equivocation by disambiguating everything. Let us rather escape them by weakening our logic so that it tolerates ambiguity; and this we can do, it turns out, by adopting some of the strictures of the relevantists.

The pessimist may be right, given his broad construal of "ambiguity" (which is perfectly appropriate in this context), in doubting that full disambiguation is feasible. But that is more of a remedy than we really need. For purposes of truth-functional logic, at least, it is good enough if we can disambiguate to the point where everything in our reasoning is true-osd only or false-osd only, even if ambiguity in sense remains. Can we do as well as that?

Certainly we often can; maybe we always can. Or maybe we sometimes can't, especially in perplexing subject matters where ambiguity is rife. I have no firm view on the question. If indeed we sometimes cannot disambiguate well enough, as the pessimist fears, then it may serve a purpose to have a partly relevant logic capable of stopping fallacies of equivocation even when equivocation is present.

There is indeed a sense in which classical logic preserves truth even in the presence of ambiguity. If an implication is classically valid, then for every unmixed disambiguation of the entire implication, in which each ambiguous constituent is disambiguated the same way throughout all the premises and the conclusion, the conclusion is true on that disambiguation if the premises are. But although this gives a clear standard of validity for ambiguous implications, the pessimist will doubt that it is a useful standard. If things are as bad as he fears, he must perforce reason from premises accepted merely as true-osd, or at best as true-osd only. If he cannot disambiguate further, how can he tell whether his premises are made true together by any unmixed disambiguation? His highest hope for his conclusion is that it will be true-osd only, or at least true-osd. But he cannot tell whether those properties carry over from his premises to his conclusion just by knowing that unmixed disambiguations of the entire implication always preserve truth.

Jaśkowski [6] put forward his "discursive logic" *inter alia* as a logic for ambiguity. His standard of validity for ambiguous implications differs from those we have considered. He requires unmixed disambiguation of each sentence that occurs as a premise or conclusion, but allows ambiguous constituent to be disambiguated differently in different sentences of an implication. If an implication is valid by Jaśkowski's standard, then if each premise is true on some unmixed disambiguation, the conclusion is so as well. The pessimist will complain that this standard too is useless, though better than the classical one. He knows that an implication satisfies Jaśkowski's standard; he accepts each premise as true-osd, or even as true-osd only; but he still cannot tell whether to accept the conclusion as true-osd, unless he can tell whether the disambiguations that make his premises true are mixed or unmixed. But how is he to tell that if he cannot disambiguate the premises?[1]

REFERENCES

[1] Alan Ross Anderson and Nuel D. Belnap, Jr., "Tautological Entailments," *Philosophical Studies* 13(1962): 9-24.

[2] Nuel D. Belnap, Jr., "How a Computer Should Think," in G. Ryle, *Contemporary Aspects of Philosophy* (Oriel Press, 1977).

(3] J. Michael Dunn, "Intuitive Semantics for First-Degree Entailments and 'Coupled Trees'," *Philosophical Studies* 29(1976): 149-68.

[4] ———, "A Kripke-Style Semantics for R-Mingle Using a Binary Accessibility Relation," *Studia Logica* 35(1976): 163-72.

[5] ———, "A Theorem in 3-Valued Model Theory with Connections to Number Theory, Type Theory, and Relevant Logic," *Studia Logica* 38(1979): 149-69.

[6] Stanisław Jaśkowski, "Propositional Calculus for Contradictory Deductive Systems," *Studia Logica* 24(1969): 143-57.

[7] D. C. Makinson, *Topics in Modern Logic* (Methuen, 1973).

[8] Robert K. Meyer, "A Boolean-Valued Semantics for R," Research Paper No. 4 of the Logic Group, Department of Philosophy, Research School of Social Sciences, Australian National University.

[9] Charles Pinter, "The Logic of Inherent Ambiguity," in A. I. Arruda, N.C.A. Da Costa, and A.M. Sette, *Proceeding of the Third Brazilian Conference on Mathematical Logic* (Sociedade Brasileira de Lógica and Universidade de São Paulo, 1980).

[10] Graham Priest, "The Logic of Paradox," *Journal of Philosophical Logic* 8(1979): 219-41.

[11] Nicholas Rescher and Robert Brandom, *The Logic of Inconsistency* (Blackwell, 1980).

[12] Richard Routley, "Ultralogic as Universal?" *Relevance Logic Newsletter* 2(1977): 50-90 and 138-75; reprinted in Routley, *Exploring Meinong's Jungle and Beyond* (Australian National University, 1980).

[13] ———, "Dialectical Logic, Semantics and Metamathematics," *Erkenntnis* 14(1979): 301-31.

[14] ——— and Valerie Routley, "The Semantics of First Degree Entailment," *Noûs* 6(1972): 335-59.

[15] P. K. Schotch and R. E. Jennings, "Inference and Necessity," *Journal of Philosophical Logic* 9(1980): 327-40.

NOTES

[1] I thank John Burgess, Max Cresswell, and J. Michael Dunn for helpful discussions.