I am sympathetic to the work in paraconsistent logics that Priest describes, but I think it should be viewed more in the form of an interesting alternative rather than a definitive improvement on what has been done before. Here I respond partly in defense of my position, but also partly in an attempt to assess the picture in terms of these two approaches that in some respects have much in common. I will present my comments in five pieces: general ideas about paradox and commonsense; metalanguage issues; an apparent disadvantage of paraconsistent logics; an apparent advantage of paraconsistent logics; and a related line of research.

Throughout I use $T$ as a truth predicate and $L$ as a liar sentence: $L$ is provably equivalent to the denial of its truth: $(L \iff \neg TL)$. I also take $Fa$ as an abbreviation for $T\neg a$.

1. General Ideas about Paradox and Commonsense

The issue can be seen in terms of a choice between the $T$-scheme and excluded middle ($T \lor F$: every wff is either true or false) on the one hand, and $ex contradictione quodlibet$ and what we might call excluded straddle\(^2\) (not $T \land F$: no wff can be both true and false) on the other. Both excluded middle and excluded straddle (as well as the $T$-scheme) are part of naive pre-paradoxical intuition. But together they produce a seemingly intolerable conflict. For in the

\(^1\)I hasten to remind the reader that "my" approach was really started by Gilmore [5] and Kripke [8] and then pursued by Feferman [4] and myself [12]. Also, many of these matters have been hotly debated in the literature for the past several years, especially in the Journal of Philosophy and the Journal of Philosophical Logic, and a number of Priest's criticisms have been made before. See Martin [9], Reinhardt [14, 15], and Goodman [7] for a glimpse of some positions.

\(^2\)This terminology was suggested to me by Mark Fulk.
case of the liar $L$ (in most accounts including both Priest’s and mine), the wffs $\neg TL$ and $\neg FL$ are provable. But excluded middle then gives $FL$ and $TL$, so that excluded straddle is violated. What is to be done? There are two obvious alternatives. Priest accepts excluded middle and rejects excluded straddle (he embraces the possibility of $T \& F$); to get away with this he also must judiciously weaken the logic (e.g., by rejecting modus ponens) so as to diffuse the derivation of all wffs from a contradiction (ex contradicitione quodlibet). The alternative is to retain excluded straddle but reject excluded middle and the $T$-scheme (in favor of a weaker version, the $T^*$-scheme); this characterizes the Gilmore–Kripke approach.\textsuperscript{3} Feferman and I follow suit in the sense that, although our logics are classical first-order ones, the truth predicate $T$ that we employ does not have the property that, for every wff $\alpha$, $Ta \lor T\neg \alpha$.$\textsuperscript{4}$ It is important to note that for many wffs $\alpha$, however, $Ta \lor T\neg \alpha$ is provable in these logics; it simply is not so for “ungrounded” wffs (such as the liar $L$).

The question then is which approach is more palatable. Priest seems to think that it is quite clear that Tarski’s $T$-scheme is necessarily part of the very meaning of truth, and with it also accepts excluded middle. I feel that the $T$-scheme and excluded middle are of course part of pre-paradoxical intuition about truth but must fall by the wayside on closer inspection. This is not a matter where there is immediate clarity; however, I have come to take a view much like that given by Reinhardt [14, 15] (which he attributes to Gödel [6]): that some wffs are plainly or fully meaningful (true or false), some are plainly jibberish, and some are partly meaningful while yet being neither true nor false. The liar fits into the third category; it is meaningful enough for us to fit it into a cogent logical argument and get something new out (e.g., that $L$ can be neither true nor false!).

Thus there are categories or degrees of meaningfulness, something which pre-paradoxical intuition did not grasp, and this is perhaps the positive lesson of the liar. The $T$-scheme then fails, because $L$ is meaningful enough to allow us to verify it ($L$ is a theorem) yet it is not fully meaningful (hence not true). That is, we have $L$ but not $TL$. Since $L$ is a strange bird anyway, I do not find it upsetting to my intuitions that $L$ disobeys the $T$-scheme. Another way to say this is that when we tag a wff $\alpha$ as “true,” we are relying on an unspoken principle, which I call the normal-order or priority principle, to the effect that $\alpha$

\textsuperscript{3}But one must not that $\alpha \lor \neg \alpha$ is retained; it is $Ta \lor Fa$ that is in general lost. This is tied to the fact that $\alpha$ and $Ta$ are not as tightly linked in the $T^*$-scheme as in the $T$-scheme. Gilmore worked in set theory, but the issues are much the same, as Feferman so aptly demonstrates. Also, yet another school of thought, Brouwerian intuitionism, rejects $T \lor F$ in general, but in a very different way.

\textsuperscript{4}Although of course it does have $Ta \lor \neg Ta$. This approach then critically distinguishes between $T\neg \alpha$ and $\neg T\alpha$. The former, $T\neg \alpha$, is identified with $Fa$. Thus there remains a kind of “truth-gap” in these theories, although the gap is expressible, even provable, in the form $\neg Ta$, for some wffs such as the liar $L$. That is, we prove both $\neg TL$ and $\neg FL$ which formally reveals the gap for $L$. 
expresses a proposition whose import is independent of the linguistic decision to call $\alpha$ true. In peculiar cases, such as the liar, our experience leads us astray, because some wffs, surprisingly, twist back and lean on their own truth-tagging! They are not totally meaningless, of course, for they at least tell us enough to allow us to make arguments based on them as we just saw. To argue that these twisted wffs “should” still obey ordinary truth-tagging conventions I think gives too much weight to naive pre-paradoxical experience.

I will style the Gilmore–Kripke approach as “$\neg(T \lor F)$” (it rejects excluded middle in general) and Priest’s alternative as “$T \& F$” (it embraces $T \& F$, thereby rejecting excluded straddle in general). Both the approaches have a very similar underlying spirit, in that each grants the existence of certain peculiar wffs that are ambiguous. Whether one prefers to regard these wffs as having no truth value or both truth values is a subtlety whose merits remain to be argued. Yet Priest seems to think that the case is clear for $T \& F$ and against $\neg(T \lor F)$. For me, it is far easier to conceive of meaningless statements (or merely partially-meaningful statements) that are neither (fully) true nor (fully) false, than to imagine a statement being both fully true and fully false at the same time. However, this is a matter to be determined on the basis of broad technical experience with many examples.

Moreover, Priest seems to vacillate between two positions: (i) that the $T$-scheme is appropriate for “an adequate representation of cognitive reasoning,” and (ii) that it “must hold for the truth predicate, that it, indeed, characterizes truth” in general. These are very different claims which he does not distinguish. Regarding claim (ii), taking $T \& F$ as a degree of meaningfulness somewhat less than either $T$ or $F$ alone, there might be room for a philosophical and even technical meeting of positions here. As for (i), see my remarks in Sections 3 and 5.

2. Metalanguage Issues

The metalanguage issue Priest raises I think distorts my claims. Indeed I had sought a language in which an agent could carry on reasoning about the truth of its own utterances. This can be done with the $T^*$-scheme in the $\neg(T \lor F)$ approach. Whether one regards $T^*$-scheme as a means for reasoning about truth, depends on one’s understanding of the meaning of the word “truth.” Note that it is the provability of $L$ that leads us to see that it is (Tarski) “true” in a model. But the provability of $L$ can be seen by the agent, via a Gödel Beweis predicate. In [13] there is some discussion of this kind of reasoning. Moreover, if we start with a language of set theory then we can describe the

---

5I make a similar point in [11]: “... the truth of a sentence can be identified with the meaning of that sentence precisely when that meaning doesn’t depend on whether we designate it as true.” (p. 25).
entire Gilmore–Kripke model-building apparatus and prove (in the very same
theory) that \( L \) comes out Tarski-satisfied there.

Now, the Tarskian definition of satisfaction of a sentence \( \alpha \) with respect to
an interpretation \( \mathcal{I} \) does in a sense provide a version of \( T \lor F \), in that \( \alpha \) is
definitely either satisfied or not satisfied ("true" or "false") in \( \mathcal{I} \). However,
such an argument is misleading in the present context. This Tarskian "semantics"
does not at all define the notion of truth. This may seem surprising, for it
often is presented as such. But in fact it presupposes that all the predicate
letters (and hence atomic wffs) are already interpreted (as "true" or "false" on
domain elements) and then the ramifications of this for more general wffs are
drawn out by Tarski. Whether or not a given atomic wff is true, is not
addressed at all by this, nor even what it means for an atomic wff to be true.
Tarskian "truth" is used in my approach only as a means to prove consistency,
not as a philosophical basis for truth.

What is more, in the present debate truth itself is formalized via a predicate
letter! So comparing formal properties of \( T \) in the \( \neg(T \lor F) \) approach with the
standard Tarskian semantics as the final arbiter, seems simply irrelevant, or
circular at best. Yet this, to my thinking, is the essence of the claim that a truth
predicate should behave according to Tarskian semantics. Of course, as noted
above, should one wish to do so, the Tarskian definition itself can be
formalized within a first-order (set) theory, although this would not bear on
what "true" means, for the reasons I have given.

3. An Apparent Disadvantage of Paraconsistent Logics

The meaning of negation in paraconsistent logic is peculiar, for \( L \) is both \( T \) and
\( F \), so also \( \neg L \) is \( T \) and \( F \). What does it mean, to have both \( L \) and \( \neg L \)? It
seems that negation takes on a highly unfamiliar sense, for \( \neg L \) customarily is
used to deny that \( L \) is the case, and yet here we have it and are denying we
have it! What is its relation to ordinary reasoning? This can be dramatized by
considering examples.

Consider a rule (a variant on one Priest gives) that "books with any true
sentences are to be destroyed, and those with only false sentences are to be
read (and not destroyed)." Then under Priest's approach a book with only the
liar sentence \( L \) (and armed with suitable cases of modus ponens) will put the
reasoner in a quandary, requiring both destruction and non-destruction. Yet
surely this flies in the face of naive intuition.\(^6\) The \( \neg(T \lor F) \) approach suggests
such a book would be viewed as satisfying neither condition, and hence no
action is required.

\(^6\) Priest might reply that in this case it is the rule of action that is at fault, that does not seem
problematic, until we get into the details. Fine, but this kind of argument is also the basis for
criticism of the \( T \)-scheme, which is a kind of rule for the use of \( T \), and it seems ok until one looks
closer. This is what has motivated people to try things like the \( \neg(T \lor F) \) approach of the
\( T^\bullet \)-scheme.
The particular book example as given by Priest, I do not find convincing in naive intuitive terms. He envisions a book with contradictory wffs. A person who notes the possibility of a twisted (self-referential) wff may very well decide not to destroy the book, with the thought that some wffs are indeed so twisted as to be neither clearly true nor clearly false, but simply ambiguous. This is in the spirit of the \( \neg (T \lor F) \) approach. Alternatively, if the command (to destroy books with truth) is interpreted to mean books with a sentence that is not false (i.e., a sentence whose negation is not true) then there is no ambiguity in the \( \neg (T \lor F) \) approach, since for every wff \( x \), either \( x \) or \( \neg x \) will be not false, that is, we have \( \neg \text{True}(x) \lor \neg \text{True}(\neg x) \). So given that a book contains a contradiction \( x \& \neg x \), if follows that the book must be destroyed.

Similarly for truth-teller stories. If \( X \) told the truth, then he is human, but if he lied then he is not human. So he is and is not human. Fine, a contradiction, which Priest informs us can do no harm. No harm? Well, it certainly harms the course of reasoning, to the extent that the poor detective cannot tell whether to arrest \( X \) (for perjury) or not! Perhaps he is both to arrest and not to arrest? An interesting commonsense world Priest has laid out for us!

Again, Priest’s discussion of the Id–Od example I think misses the point of the \( \neg (T \lor F) \) approach. To be specific, if what Id says is not true, and if Od said just that, it need NOT be the case that Od spoke truely, despite naive intuition. But if we agree that speaking non-falsely (rather than truthfully) entails being human, then \( \neg (T \lor F) \) gives the same result as \( T \& F \).

4. An Apparent Advantage of Paraconsistent Logics

A possible deficiency in the \( \neg (T \lor F) \) approach that Priest does not mention, but that can be seen as an advantage for paraconsistent logics, is that of the need to single out a distinction between True and not-False before writing axioms for a problem. This surfaces in the Id–Od and book examples above. But the approach Priest advocates seems to avoid any such distinction, which certainly makes the formalization process easier.

5. A Related Line of Research

There is much to be said for allowing contradictions to appear in commonsense reasoning. After all, human reasoners often encounter contradictions in their own beliefs, and do not then proceed to believe everything else! However, this is not necessarily a point in favor of the paraconsistent logics Priest discusses. For when a (direct) contradiction is found in our reasoning, we tend to notice that fact and take corrective action, such as temporarily suspending belief in one or both conflicting beliefs. Priest alludes to this, but it is not part of paraconsistent logic as such, and other work (mentioned below) suggests that even more massive changes in logic are called for to address this problem.

But see [10].
In particular, detecting contradictions is generally only semidecidable, and this would appear also to be the case for paraconsistent logics. This means in general that computation cannot proceed, if it depends on first finding any present contradictions. It would seem that a major revision in our idea of logic for (effective) commonsense reasoning is called for. In [1–3] the theme of permitting contradictions and the potential for their revision in an effective way is studied. While these "step-logics" are not strictly paraconsistent, they do in a sense embrace a failure of ex contradictione quodlibet measured over finite time steps. That is, conclusions are seen as emanating over time, and contradictions can be detected before they lead to many nonsensical conclusions.

ACKNOWLEDGMENT
I would like to thank Mark Fulk for helpful discussions, and the University of Rochester for a stimulating environment during my sabbatical stay there as this paper was finished. This research was supported in part by a grant from the Army Research Office (DAAL.03-88-K0087).

REFERENCES
3. Elgot-Drapkin, J. and Perlis, D., Reasoning situated in time I: Basic concepts (Submitted).

Received October 1988; revised version received November 1988
Truth definition: The truth about something is all the facts about it, rather than things that are imagined | Meaning, pronunciation, translations and examples. Example sentences containing 'truth'. These examples have been automatically selected and may contain sensitive content. Read more.| I'm not as adept as you and Tarja at twisting the truth to placate my honour, Lord Wolfblade.