On Priest on nonmonotonic and inductive logic

Greg Restall
School of Historical and Philosophical Studies
The University of Melbourne
Parkville, 3010, Australia

restall@unimelb.edu.au
http://consequently.org/

Abstract: Graham Priest defends the use of a nonmonotonic logic, *LPm*, in his analysis of reasoning in the face of true contradictions, such as those arising from the paradoxes of self-reference. In the course of defending this choice of logic in the face of the criticism that this logic is not truth preserving, Priest argued (2012) that requirement is too much to ask: since *LPm* is a nonmonotonic logic, it necessarily fails to preserve truth. In this paper, I show that this assumption is incorrect, and I explain why nonmonotonic logics can nonetheless be truth preserving. Finally, I diagnose Priest's error, to explain when nonmonotonic logics do indeed fail to preserve truth.

# # #

Non-classical logics are not *classical*. Sometimes this fact seems like a feature: paraconsistent logics like Priest's *LP* reject disjunctive syllogism and *ex contradictione quodlibet* and so, they give us new and fruitful ways to deal with semantic paradoxes unavailable to proponents of classical logic. However, sometimes this fact seems like a bug: there are times we want to endorse those particular rules of proof, not reject them. Priest's favoured way out of this tension is to adopt a nonmonotonic logic, *LPm*. According *LPm*, inference steps such as disjunctive syllogism—from $p \vee q$ and $\neg p$ to q—may be valid, while becoming invalid in the presence of extra premises: in particular, premises which are inconsistent.

The precise details of how *LPm* manages to be nonmonotonic are not important to us. A sketch will suffice to explain what is going on. Models for *LP* and *LPm* assign truth (1) or falsity (0) or possibly *both* to each atomic proposition. An argument is *LP* valid if every model that assigns the premises to be true (perhaps some premises are false too, perhaps they are not) also renders the conclusion true. For *LPm* we relax this condition. We need not check every model—in particular, we don't need to check the models that assign both 1 and 0 to many different propositions. We check models that are as inconsistent as they need to be to make the premises true. (These are the *minimally inconsistent* models, and that is where the "*m*" comes from in "*LPm*".) If in every such model where the premises are true, so is the conclusion, then the argument is *LPm* valid. So, disjunctive syllogism in the shape of the argument from $p \vee q$ and $\neg p$ to q is *LPm*-valid, since there are completely consistent models in which the premises are true (these are the consistent model in which p is false but q is true), and in these models, the conclusion q is indeed true. We can disregard models in which p is both true and false, as we never need p to be inconsistent to make the premises true. If we add the premise $p \wedge \neg p$, then the models that make the premises true have to be inconsistent about p. The models in which p is both true and false and q is false only is no better and no worse than models in which p is both true and false and q is true only. But the models like this in which q is false are counterexamples to the argument—they make the premises true and the conclusion false, so adding the inconsistency of p as an extra premise renders the new argument *invalid* in *LPm*.

Much of Priest's work in paraconsistent logic uses the relatively traditional, monotonic logic *LP* rather than the stronger nonmonotonic logic *LPm*. Many theories which are trivial in the context of classical logic (in the sense that there are no models at all, and everything follows from the axioms of the theory) are non-trivial in *LP*. This gives rise to the concern that since *LPm* is stronger than *LP*, some of the theories which are non-trivial in *LP* may be trivial when viewed through the lens of *LPm*. Priest proves (2006, page 226) that Reassurance indeed holds for *LPm*.

In a recent paper, Beall (2012) argues that this Reassurance theorem is not enough to be genuinely reassuring. He claims that we should want what he calls *General Reassurance*: if the *LP* consequences of some set of premises are true, so are the *LPm* consequences of that set. In reply to Beall, Priest (2012) argues that General Reassurance is too much to ask of *LPm* of or any nonmonotonic logic. He writes (2012, page 740):

> *General Reassurance*, however, is too much to ask. *LPm* is a nonmonotonic (aka inductive) logic. And it is precisely the definition of such logics that they may lead us from truth to untruth. The point is as old as Hume ('The sun has risen every day so far. So the sun will rise tomorrow.') and as new as that much over-worked member of the spheniscidae ('Tweety is a bird. So Tweety flies.') If they did not have this property, these logics would be deductive logics, which they are not. This is not a bug of such logics; it is a feature. Such logics do not preserve truth, by definition.

I will not attempt to adjudicate the disagreement between Priest and Beall on the virtues of General Reassurance—this would require settling what the consequence relation of *LPm* is *for*, and that is beyond the scope of this note. Here, I have a simpler point to make. Priest's characterisation of the relationship between non-deductive, non-truth-preserving logics and nonmonotonic logics is mistaken, and I will explain why, giving examples of truth preserving nonmonotonic logics. Once I have presented the counterexamples to Priest's claim, I will attempt to diagnose his error, and explain why one might reasonably, but mistakenly, take it that a nonmonotonic logic is never truth preserving.

# # #

Let me be careful to define our terms:

- A consequence relation $\vDash$ is *nonmonotonic* if there are valid arguments from premises Σ to conclusion C ($\Sigma \vDash C$) such that there is some extra premise B where the argument from Σ together with B to C fails to be valid ($\Sigma, B \nvDash C$).

- A consequence relation $\vDash$ is *non-truth-preserving* (or 'inductive' or 'non-deductive') if there is some valid argument from premises Σ to conclusion C ($\Sigma \vDash C$) where each premise in Σ is true and the conclusion C is false.

Notice that these definitions use different concepts. It would be surprising for them to coincide, and in fact it is not hard to find examples of nonmonotonic but truth preserving consequence relations. (It is even easier to find monotonic logics that fail to preserve truth. Consider the 'consequence relation' for which an argument is 'valid' if whenever the premises all contain the letter '*e*' so does the conclusion. This is monotonic, but it fails to preserve truth.)

Example 1: Not all friends of paraconsistent logics are dialetheists. You can reject the inference from a contradiction to an arbitrary conclusion, without taking any contradictions to be *true*. Nonetheless, there are *models* in which contradictions are true. Those models represent different ways that things can't be (Restall 1997). For a paraconsistentist who takes contradictions to be semantically distinct (and to have different consequences) but nonetheless all impossible, the logic *LPm* is truth preserving. If all possible worlds are consistent and complete, then any *LPm* -valid argument leads from truths only to other truths, and necessarily so, for any world there is a consistent *LPm* model assigning 1 to each truth and 0 to each falsehood of the language, so if the premises of an *LPm* -valid argument are true in some possible world, the conclusion must be true too, since the model appropriate to that world is as minimally inconsistent as you can get—it is actually *consistent*.

So, the logic is now truth preserving, but it remains nonmonotonic. While disjunctive syllogism is *LPm* valid, the addition of the inconsistent premise renders the argument invalid. The model that delivers the invalidity is not a *possibility* for the non-dialethic paraconsistentist. It represents a way that things cannot be.

Now, Priest is a dialetheist, so while he should agree that the nondialethic paraconsistentist can take *LPm* to be a truth preserving nonmonotonic logics (and so, this is enough to show that being nonmonotonic alone is not enough to be non-truth-preserving), dialetheists cannot use that example for themselves. However, it's easy enough to construct examples that are dialethically acceptable to make the same point. If the actual world is inconsistent about some things but not others—say, for the actual world we require inconsistency over *some* part of our vocabulary, L_1 and not the rest, L_2—then take the 'baseline' for inconsistency to be models that are inconsistent only in L_1 but not in L_2, and grade models which allow for more or less inconsistency in the L_2 vocabulary. The resulting consequence relation is now truth preserving but still nonmonotonic.

Example 2: Consider models for counterfactuals that use a similarity relation on worlds, in the style of Lewis or Stalnaker. A conditional $A > B$ is true at world w if in the worlds worlds *most similar* to w where A is true, so is B. Let's say that an argument from premises Σ to conclusion C is *Cf*-valid if the conditional $\bigwedge\Sigma > C$ is true. These conditionals are famously nonmonotonic. (The closest worlds where I have a cup of coffee before 7am are not the closest worlds where I have a cup of coffee with added arsenic before 7am.) However, *Cf*-valid arguments are truth preserving, given the plausible assumption (shared by Lewis, Stalnaker and others who take this approach to counterfactual conditionals) that a world w is one of the closest worlds to itself. Here is why. Suppose the argument from Σ to C is *Cf*-valid, and that each sentence in Σ is true. We want to show that C is true too. Since the argument is *Cf*-valid, at the actual world, the conditional $\bigwedge\Sigma > C$ is true. Since the actual world is one of the closest worlds to itself, and since $\bigwedge\Sigma$ is true at the actual world, C is true there too, as desired. This (contingent, non-formal) 'logic' of conditional consequence gives us another example of a nonmonotonic but truth preserving consequence relation.

Examples 3, 4, …: We can make arbitrarily more examples of nonmonotonic and truth preserving logics using a simple template. Given a monotonic consequence relation $\vDash$ defined in terms of truth preservation with some class M of models, in which we have settled in advance that the actual world represented by some model in a subclass W of M. (In *Example 1*, W is the class of consistent *LP* models. In *Example 2*, it is the class containing all worlds most similar to the actual world.) We enrich our interpretation with a well-founded preorder relation $\sqsubseteq$, according to which each member of W is minimal according to that relation—for each $w \in W$ there is no $v \in M$ where $v \sqsubseteq w$ but $w \not\sqsubseteq v$. (The well-foundedness condition ensures that every nonempty subset of M has elements that are $\sqsubseteq$-minimal in M.) Define the non-monotonic consequence relation $\vDash^*$ by setting $\Sigma \vDash^* C$ if the $\sqsubseteq$-least models in which each element of Σ is true also make C true.

The consequence relation $\vDash^*$ is truth preserving by design. If $\Sigma \vDash^* C$ and the members of Σ are true, then there is some model in W (the model of the actual world, which makes true all and only the true sentences), in which the members of Σ hold. Since this model is in W it is minimal with respect to $\sqsubseteq$ and since $\Sigma \vDash^* C$, then C holds at that model too, and so, it is true.

We need to do a little more work to show that $\vDash^*$ is not monotonic. For that we need some information about the language and the class M of models and its subset W. If there is an argument in our language from Σ to C that has no counterexamples among *worlds* (in W) but has some counterexample in a model m outside W, then if we have some sentence B_m true at m but not true at any world in W, our argument will be a counterexample to monotonicity—we have $\Sigma \vDash^* C$ but we don't have $\Sigma, B_m \vDash^* C$, since there is some model $\sqsubseteq$-minimal among models in which Σ, B_m are true and C is untrue, since m is one such model, the set of all such models, being non-empty, must have a $\sqsubseteq$-minimal member.

This technique is general, and it shows that there are many different ways to construct nonmonotonic but truth preserving consequence relations. Priest was mistaken to identify nonmonotonicity with failure to preserve truth.

# # #

That demonstrates the scope of the error. It is another thing to diagnose it. Why might you think that there is a connection between nonmonotonicity and the failure to preserve truth? Consider Priest's motivating examples of nonmonotonic inferences: 'The sun has risen every day so far. So the sun will rise tomorrow', 'Tweety is a bird. So Tweety flies.' If you take those inference steps to be unrestrictedly valid in the target sense of validity, then we surely have counterexamples to truth preservation. *Some* nonmonotonic consequence relations are not truth preserving. Which ones fail to be truth preserving? Is there a deeper connection between nonmonotonicity and failure to preserve truth?

Here is one possible connection. Suppose the consequence relation $\vDash^*$ satisfies the following conditions:

1. $\vDash^*$ invalid arguments are witnessed by models. If $\Sigma \nvDash^* A$ then there is some m where each statement in Σ holds in m but A does not hold in m.
2. The models m used in (1) are all *possibilities*. If m is a model, and A is true in m then A is *possible*.
3. $\vDash^*$-validity is not world-relative. If an argument is valid, then had things been otherwise, it still would have been valid.

Under these three conditions, any failure of monotonicity gives rise to a failure of truth preservation. (These are sufficient conditions, not necessary conditions. There are many *other* conditions under which nonmonotonic logics may fail to preserve truth.) Take a failure of monotonicity, where $\Sigma \vDash^* C$ but $\Sigma, B \nvDash^* C$. By (1) there is some model m which is a counterexample to the argument from Σ, B to C. If the world is like m, then the argument from Σ to C, though valid according to $\vDash^*$, would not be truth preserving, since though each member of Σ is true at m, the conclusion C is not. By (2), this is a *possible* failure of truth preservation of the argument from Σ to C (since what is true at m is indeed possible), and by (3), what *is* valid (the argument from Σ to C) still would have been valid at that circumstance, so this is indeed a circumstance where a valid argument has a counterexample—it is a failure of truth preservation.

These three conditions are plausible constraints on certain kinds of consequence relations, and I conjecture that Priest endorses all three (for an appropriate way of understanding the class of models in question). If so, this explains why Priest would be reasonable to make the step from nonmonotonicity to failure to preserve truth, despite the counterexamples we have seen.

Why does this argument not work for the examples of truth preserving nonmonotonic logics given in the previous section? For the non-dialethic paraconsistentist, (2) fails. Inconsistent models are not all *possibilities*. The non-dialethic paraconsistentist agrees that there is a *model* in which a contradiction $p \wedge \neg p$ is true while an arbitrary q is not (which is a witness to the failure of the argument from $p \wedge \neg p$ to q), but such a model is a way that things cannot be, not a way that things can. For counterfactual consequence, the worlds are each possibilities, but (3) fails. Consequence is contingent and world-relative. This argument breaks down because although a world might be a counterexample to the argument from Σ, B to C, it doesn't follow that if the world was like *that*, then we would have a counterexample to the valid argument Σ to C, because from the point of view of *that* world, the argument from Σ to C *is* Cf-valid.

So, nonmonotonic consequence relations are closely connected with failures of truth preservation, but that connection is not identity. Understanding *this* connection is an important aspect of understanding the many different connections between consequence relations, possibility and truth.[1]

References

Beall, Jc. 2012. "Why Priest's Reassurance is not so Reassuring." *Analysis* 72:517–525.

[1] Thanks to Jc Beall, Graham Priest and Shawn Standefer for comments on this paper.

Priest, Graham, 2006. *In Contradiction*, Second Edition. Oxford: Oxford University Press.

Priest, Graham. 2012. "The Sun may Not, Indeed, Rise Tomorrow: a reply to Beall." *Analysis* 72:739–741.

Restall, Greg. 1997. "Ways Things Can't Be." *Notre Dame Journal of Formal Logic* 38:583–596.