A Worked Countermodel Question

Thomas Forster

September 27, 2006

Challenge: Find a countermodel for $((A \to B) \to A) \to A$.

The first thing to notice is that this formula is a classical (truth-table) tautology. This means that any countermodel for it must contain more than one world.

Now a model $\mathfrak{M}$ satisfies a formula ψ iff the root world of $\mathfrak{M}$ believes ψ: that what it is for a model to satisfy ψ. Definition! (Sometimes in the literature the root world is called the "designated" or "real" world).

Let me write 'W_0' for the root world, and let me also write '$W \models \psi$' to mean that the world W believes ψ, and $W \not\models \psi$ to mean that W does not believe ψ. (I'm using these mathematical symbols not merely to annoy you—tho' of course it will—but to make it easier to do back-of-the-envelope calculations of the kind you are about to see!)

So we know that $W_0 \not\models ((A \to B) \to A) \to A$.

Now the definition of $W \models X \to Y$ is

$$(\forall W' \geq W)(W' \models X \ \to \ W' \models Y) \tag{1}$$

("W believes $A \to B$ iff every world visible from W that believes X also believes Y"—you can see why I prefer the symbols!)

So since

$$W_0 \not\models ((A \to B) \to A) \to A$$

we know that there must be a $W' \geq W_0$ which believes $((A \to B) \to A$ but does not believe A. (In symbols: $(\exists W' \geq W_0)(W' \models ((A \to B) \to A) \ \wedge \ W' \not\models A$.) Remember too that in the metalanguage we are allowed to exploit the equivalence of $\neg\forall$ with $\exists\neg$. Now every world can see itself, so might this W' happen to be W_0 itself? No harm in trying...

So, on the assumption that this W' that we need is W_0 itself, we have:

1. $W_0 \models ((A \to B) \to A)$; and
2. $W_0 \not\models A$.

This is quite informative. Fact (1) tells us that every $W' \geq W_0$ that believes $A \to B$ also believes A. Now one of those W' is W_0 itself (every world can see itself remember that $\geq$ is reflexive). Put this together with fact (2) which says that W_0 does not believe A, and we know at once that W_0 cannot believe $A \to B$. How can we arrange for W_0 not to believe $A \to B$? Recall the definition (1) above of $W \models A \to B$. We have to ensure that there is a $W' \geq W_0$ that believes A but does not believe B. This W' cannot be W_0 because W_0 does not believe A. So there must be a *new* world (we always knew there would be!) visible from W_0 that believes A but does not believe B. (In symbols this is $(\exists W' \geq W_0)(W' \models A \wedge W' \not\models B)$)

So our countermodel contains two worlds W_0 and W', with $W_0 \leq W'$. $W' \models A$ but $W_0 \not\models A$, and $W' \not\models B$.

Let's check that this really works. We want

$$W_0 \not\models ((A \to B) \to A) \to A$$

We have to ensure that at least one of the worlds beyond W_0 satisfies $(A \to B) \to A$ but does not satisfy A. W_0 doesn't satisfy A so it will suffice to check that it does satisfy $(A \to B) \to A$. So we have to check (i) that if W_0 satisfies $(A \to B)$ then it also satisfies A and we have to check (ii) that if W' satisfies $(A \to B)$ then it also satisfies A. W' satisfies A so (ii) is taken care of. For (i) we have to check that W_0 does not satisfy $A \to B$. For this we need a world $\geq W_0$ that believes A abut does not believe B and W' is such a world.