Title  Counterpossibles and the Nature of Impossible Worlds

Keywords  Counterfactuals · Counterpossibles · Impossible worlds · Lewis-Stalnaker Semantics · Non-classical Logics

Abstract  Since the publication of Stalnaker’s “A Theory of Conditionals” [6] and Lewis’ Counterfactuals [4], the standard semantic account of counterfactuals has been phrased within the possible worlds framework. Roughly, according to Stalnaker and Lewis, a conditional of the form ‘If $\varphi$ had been the case, $\psi$ would have been the case’ (represented as ‘$\varphi \mathrel{\Box} \psi$’) is true just in case $\psi$ is true in the closest possible world in which $\varphi$ is true, where ‘closest’ is read as ‘most similar in relevant respects to the world of utterance.’

According to the standard Lewis-Stalnaker semantics, all counterfactuals with necessarily false antecedents (or counterpossibles) are vacuously true. That is, if $\varphi$ is true in no possible world, then $\varphi \mathrel{\Box} \psi$ is true, for any $\psi$. Many theorists have found a vacuous treatment of counterpossibles problematic, because some counterpossibles seem non-trivially true and others non-trivially false. Borrowing from Nolan [5, p. 544], compare:

(1) If Hobbes had squared the circle, sick children in the mountains of South America at the time would not have cared.

(2) If Hobbes had squared the circle, everything would have been the case.

Intuitively, (1) is non-trivially true, whereas (2) is non-trivially false. This challenges the standard Lewis-Stalnaker semantics as a general semantics for all counterfactuals.

To solve the problem of counterpossibles, a number of theorists have proposed to extend the modal space underlying the standard Lewis-Stalnaker semantics by impossible worlds, where necessary falsehoods can be true. Roughly, according to the extended Lewis-Stalnaker semantics, $\varphi \mathrel{\Box} \psi$ is true just in case $\psi$ is true in the closest world (whether possible or impossible) where $\varphi$ is true. Given this semantics, (1) will come out non-trivially true insofar as the closest (impossible) world where Hobbes squares the circle is a world where sick children in the mountains of South America at the time do not care about it. By contrast, (2) will come out non-trivially false.
since, intuitively, not everything is the case in the closest (impossible) world where Hobbes squares the circle.

Proponents of an abstractionist ontology of impossible worlds have commonly identified these with maximal, inconsistent sets of sentences, where a set \( \Gamma \) of sentences is maximal just in case, for any sentence \( \varphi \), either \( \varphi \) or \( \neg \varphi \) is a member of \( \Gamma \). Among theorists who identify impossible worlds with maximal inconsistent sets of sentences are Brogaard & Salerno [2], Goodman [3], and Vander Laan [7, 8]. However, Bjerring [1] has recently argued that the extended Lewis-Stalnaker semantics fails to deliver the correct truth values for many counterpossibles, if impossible worlds are required to be maximal. To avoid this negative result, Bjerring considers two alternative world-ontologies in which impossible worlds need not be maximal:

**Partial Modal Space** in which impossible worlds correspond to arbitrary inconsistent sets of sentences (whether maximal or partial).

**Stratified Modal Space** in which impossible worlds correspond to (maximal or partial) sets of sentences that are closed under logical consequence in some non-classical logic.

Bjerring prefers stratified modal space over partial modal space. This preference is based on a worry to the effect that partial modal space is too comprehensive. That is, if we do not impose any constraints on impossible worlds, our modal space will, according to Bjerring, end up containing worlds that we want excluded from our modal space.

I will argue that Bjerring’s worry about partial modal space is based on a conflation of two distinct conceptions of what it means for a logic to be true in a world:

\( \mathcal{L} \)-**MEMBERSHIP**: A logic \( \mathcal{L} \) is true in a world \( w \) just in case the sentence ‘\( \mathcal{L} \) is correct (true, adequate)’ is a member of \( w \).

\( \mathcal{L} \)-**CLOSURE**: A logic \( \mathcal{L} \) is true in a world \( w \) just in case \( w \) is closed under logical consequence in \( \mathcal{L} \).

Once the distinction between \( \mathcal{L} \)-**MEMBERSHIP** and \( \mathcal{L} \)-**CLOSURE** is appreciated, Bjerring’s worry about partial modal space vanishes. I go on to argue that stratified modal space, on the other hand, is too constrained and that we should base the extended Lewis-Stalnaker semantics on partial modal
space. I also suggest a closeness condition on worlds in partial modal space which ensures that the extended Lewis-Stalnaker semantics—when based on partial modal space—delivers the correct truth values for all counterfactuals that make claims about what would be the case, had some logic (whether classical or non-classical) been true.

References


