Mattias Skipper Rasmussen*

Counterpossibles and the nature of impossible worlds

Abstract: One well-known objection to the traditional Lewis-Stalnaker semantics of counterfactuals is that it delivers counterintuitive semantic verdicts for many counterpossibles (counterfactuals with necessarily false antecedents). To remedy this problem, several authors have proposed extending the set of possible worlds by *impossible* worlds at which necessary falsehoods may be true. Linguistic ersatz theorists often construe impossible worlds as maximal, inconsistent sets of sentences in some sufficiently expressive language. However, in a recent paper, Bjerring (2014) argues that the "extended" Lewis-Stalnaker semantics delivers the wrong truth-values for many counterpossibles if impossible worlds are required to be maximal. To make room for non-maximal or partial impossible worlds, Bjerring considers two alternative world-ontologies: either (i) we construe impossible worlds as arbitrary (maximal or partial) inconsistent sets of sentences, or (ii) we construe them as (maximal or partial) inconsistent sets of sentences that are closed and consistent with respect to some non-classical logic. Bjerring raises an objection against (i), and suggests that we opt for (ii). In this paper, I argue, first, that Bjerring's objection against (i) conflates two different conceptions of what it means for a logic to be true at a world. Second, I argue that (ii) imposes too strong constraints on what counts as an impossible world. I conclude that linguistic ersatzists should construe impossible worlds as arbitrary (maximal or partial) inconsistent sets of sentences.

Keywords: counterfactuals, counterpossibles, impossible worlds, Lewis-Stalnaker semantics.

1 Introduction

Since the publication of Robert Stalnaker's "A Theory of Conditionals" (1968) and David Lewis' *Counterfactuals* (1973), the standard semantic account of counterfactuals has been phrased within the "possible worlds" framework. Roughly, according to Stalnaker and Lewis, a counterfactual conditional, "if φ had been

*Corresponding author: Mattias Skipper Rasmussen, Department of Culture and Society, Aarhus University, Jens Chr. Skous vej 7, E-mail: mattias.skipper.rasmussen@post.au.dk

the case, ψ would have been the case," is true if and only if φ is true in no possible world or ψ is true in the closest (i.e. most relevantly similar[1]) possible world in which φ is true. Call this the *standard Lewis-Stalnaker semantics* of counterfactuals.

On the standard Lewis-Stalnaker semantics, all counterpossibles (counterfactuals with necessarily false antecedents) are vacuously or trivially true: if φ is true in no possible world, then the counterfactual conditional "φ implies ψ" is true, for any ψ. This is widely regarded as an embarrassment to the standard Lewis-Stalnaker semantics since many counterpossibles appear non-trivially true or false. Borrowing from Nolan (1997: 544), consider:

(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. But the standard Lewis-Stalnaker semantics deems both (1) and (2) vacuously true.

Several authors have proposed to solve the problem of counterpossibles by extending the set of possible worlds by *impossible* worlds at which necessary falsehood may be true.[2] Roughly, according to *the extended Lewis-Stalnaker semantics*, the counterfactual conditional "φ implies ψ" is true if and only if ψ is true in the closest (possible or impossible) world where φ is true. On this semantics, (1) comes 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; and (2) comes out non-trivially false insofar as not everything is the case in the closest impossible world where Hobbes squares the circle.

Linguistic ersatz theorists often construe impossible worlds as maximal, inconsistent sets of sentences in some sufficiently expressive language, where a set Γ of sentences is maximal if and only if, for any sentence φ, either φ or $\neg\varphi$ is a member of Γ.[3] However, Bjerring (2014) has recently argued that the extended Lewis-Stalnaker semantics delivers the wrong truth-values for many

---

1 See Lewis (1973, 67) for a discussion of closeness as comparative similarity. Crucially, for Lewis, comparative similarity is a vague notion, and context plays an important role in determining which similarities and differences between various scenarios are relevant in determining the relative closeness of those scenarios to the world of utterance.

2 See Brogaard and Salerno (2013), Goodman (2004), Nolan (1997; 2013), and Vander Laan (2004) for accounts of counterpossibles that make use of impossible worlds.

3 See Adams (1974) for an early linguistic construction of worlds. For the purposes of this paper, I will sidestep issues about what an appropriate world-making language should look like, or indeed whether such a language can be specified. See Jago (2013; 2014) for a recent discussion of this issue.

counterpossibles if impossible worlds are required to be maximal. To make room for non-maximal or partial impossible worlds, Bjerring considers two alternative world-ontologies: either (i) we construe impossible worlds as arbitrary (maximal or partial) inconsistent sets of sentences, or (ii) we construe them as (maximal or partial) sets of sentences that are closed and consistent with respect to some non-classical logic. Against (i), Bjerring argues that impossible worlds end up being too unconstrained or ill-behaved if we do not impose any logical structure on them. To avoid this, Bjerring suggests that we opt for (ii).

In this paper, I argue, first, that Bjerring's objection against (i) conflates two different conceptions of what it means for a logic to be true at a world. Once this conflation is recognized, Bjerring's worry about (i) can be dismissed. Second, I argue that (ii) imposes too strong constraints on what counts as an impossible world. I conclude that (i) is preferable to (ii): linguistic ersatzists should construe impossible worlds as arbitrary (maximal or partial) inconsistent sets of sentences.

Here is the plan. In §2, I refine the rough formulations of the standard and extended Lewis-Stalnaker semantics from above. §3 is the main section of the paper. In §3.1, I present Bjerring's argument against a maximality requirement on impossible worlds. In §3.2, I argue that Bjerring's objection against (i) fails. In §3.3, I argue against (ii). This, in turn, amounts to a case for (i). In §4, I conclude.

2 Preliminaries

Let us begin by refining the rough formulation of the standard Lewis-Stalnaker semantics from §1:

> **Standard Lewis-Stalnaker Semantics (SLS):** A counterfactual conditional, "if φ had been the case, ψ would have been the case," is true in a world w if and only if there is no possible φ-world or there is some possible $\{\varphi, \psi\}$-world that is closer to w than any possible $\{\varphi, \neg\psi\}$-world.

SLS favors the Lewisian analysis of counterfactuals over the Stalnakerian one as it presupposes neither the uniqueness assumption nor the limit assumption:

> **The Uniqueness Assumption**: For any world w and proposition φ, there is a unique closest world to w in which φ is true.

> **The Limit Assumption**: For any world w and proposition φ, there is a unique (possibly singleton) set of closest worlds to w in which φ is true.

The uniqueness assumption says that worlds cannot tie for closeness to the world of utterance. The limit assumption says that worlds cannot get closer and closer to the world of utterance without end; there is a limit to how close worlds can get to the world of utterance.[4]

SLS quantifies over the set of all possible worlds, or, as I will call it, *standard modal space*. This means that SLS is bound to give a vacuous treatment of counterpossibles: if no possible world verifies φ, SLS deems the counterfactual "φ implies ψ" vacuously true, for any ψ. While Stalnaker and Lewis both defended a vacuous treatment of counterpossibles, many others have found vacuism unappealing.[5] I will not enter a discussion of the merits or demerits of non-vacuism here. For the purposes of this paper, I will simply assume that an adequate semantics of counterfactuals should allow for counterpossibles to be non-vacuously true or false.

To make room for a non-vacuous treatment of counterpossibles, several authors have proposed to augment standard modal space by impossible worlds at which (logical, metaphysical, or conceptual) impossibilities may be true. By quantifying over the set of all possible and impossible worlds—call it *extended modal space*—we get the following modified version of the Lewis-Stalnaker semantics:

> **Extended Lewis-Stalnaker Semantics (ELS)**: A counterfactual conditional, "if φ had been the case, ψ would have been the case," is true in w if and only if there is some (possible or impossible) $\{\varphi, \psi\}$-world that is closer to w than any (possible or impossible) $\{\varphi, \neg\psi\}$-world.

To illustrate how ELS is supposed to give a non-vacuous account of counter-possibles, consider the following conditionals:

(3) If intuitionistic logic were true, conjunction elimination would be invalid.
(4) If paraconsistent logic were true, the principle of explosion would be invalid.

Assuming that classical logic is true, intuitionistic logic and paraconsistent logic are both necessarily false; so (3) and (4) are counterpossible conditionals. Thus, SLS deems both (3) and (4) vacuously true. By contrast, ELS (rightly) deems (3)

---

4 See Lewis (1973, 77–83) for a discussion and rejection of the uniqueness and limit assumptions. It is worth noticing that the uniqueness assumption is strictly stronger than the limit assumption: the uniqueness assumption implies the limit assumption, but not *vice versa*.
5 See Nolan (1997) and Brogaard and Salerno (2013) for discussions of this issue.

non-vacuously false insofar as conjunction elimination is not invalid in the closest impossible world in which intuitionistic logic is true. Similarly, ELS (rightly) deems (4) non-trivially true insofar as the principle of explosion does not hold in the closest impossible world in which paraconsistent logic is true.[6]

3 The ontology of impossible worlds

Linguistic ersatz theorists often construe possible worlds as maximal, consistent sets of sentences, and *im*possible worlds as maximal, *in*consistent sets of sentences.[7] However, as mentioned in §1, Bjerring (2014) has argued that ELS delivers the wrong truth-values for many counterpossibles if impossible worlds are required to be maximal. Here follows (a slightly simplified version of) Bjerring's argument.

3.1 Bjerring against maximality

Bjerring's argument against a maximality requirement on impossible worlds is based on a proof to the effect that any maximal, inconsistent set of sentences must contain either a *contradiction*, a *conjunction fallacy*, or a *negated conjunction fallacy*:[8]

Contradiction (C): $\{\varphi, \neg\varphi\}$,
Conjunction Fallacy (CF): $\{\varphi \wedge \psi, \neg\varphi\}$, $\{\varphi \wedge \psi, \neg\psi\}$, $\{\varphi \wedge \psi, \neg\varphi, \neg\psi\}$,
Negated Conjunction Fallacy (NCF): $\{\neg(\varphi \wedge \psi), \varphi, \psi\}$.

Bearing this result in mind, Bjerring asks us to consider the following counterpossible:

(5)   If intuitionistic logic were true and $\varphi \wedge \psi$ true, then φ and ψ would be true.

---

6 Obviously, these semantic verdicts ultimately depend on our closeness metric on extended modal space. I will not enter into a discussion of principles governing relative closeness among worlds in extended modal space as my concern in this paper is with the ontology of impossible worlds. See Brogaard and Salerno (2013) and Goodman (2004) for discussions of closeness in an impossible worlds setting.
7 Brogaard and Salerno (2013), Goodman (2004), Vander Laan (1997; 2004), among others, adopt this conception of impossible worlds.
8 See Bjerring (2014, 351–352) for the proof. Nothing of importance hinges on the choice of logical connectives; any adequate set of connectives may be substituted for negation and conjunction. The resulting set of inconsistencies will vary in accordance with the choice of connectives.

Since conjunction elimination is valid in intuitionistic logic, ELS ought to deem (5) true.

Next, consider the following five counterpossibles all of which have the same antecedent as (5):

(6) (a) If intuitionistic logic were true and $\varphi \wedge \psi$ true, then χ and $\neg\chi$ would be true (for some χ).
   (b) If intuitionistic logic were true and $\varphi \wedge \psi$ true, then $\chi \wedge \omega$ and $\neg\chi$ would be true (for some χ and ω).
   (c) If intuitionistic logic were true and $\varphi \wedge \psi$ true, then $\chi \wedge \omega$ and $\neg\omega$ would be true (for some χ and ω).
   (d) If intuitionistic logic were true and $\varphi \wedge \psi$ true, then $\chi \wedge \omega$, $\neg\chi$, and $\neg\omega$ would be true (for some χ and ω).
   (e) If intuitionistic logic were true and $\varphi \wedge \psi$ true, then $\neg(\chi \wedge \omega)$, χ, and ω would be true (for some χ and ω).

Since intuitionistic logic does not license C-, CF-, or NCF-inconsistencies, ELS ought to deem all of (6a)–(6e) false. But if impossible worlds are required to be maximal, ELS must deem at least one of (6a)–(6e) true, if it deems (5) true. To see why, suppose, for the sake of simplicity, the uniqueness assumption.[9] If ELS deems (5) true, it follows that φ and ψ are true in the closest world w where intuitionistic logic is true and $\varphi \wedge \psi$ true. Since intuitionistic logic is true in w, w is an impossible world. So if impossible worlds are required to be maximal, w will contain a C-, CF-, or NCF-inconsistency. This means that at least *one* of the consequents in (6a)–(6e) will be true in w. So if ELS deems (5) true as it should, it will wrongly deem at least one of (6a)–(6e) true. The upshot is that ELS fails as a general semantics of counterfactuals if impossible worlds are required to be maximal.[10]

3.2 Relaxing the maximality requirement

In light of this negative result, Bjerring considers two alternative world-ontologies that allow for impossible worlds to be partial:

---

9 Bjerring (2014, 335–351) generalizes his argument to the case where the limit assumption (and thereby the uniqueness assumption) is not presupposed.
10 It is worth noticing that Bjerring's argument against maximality cannot be dismissed by adopting a broad notion of "inconsistency" that covers metaphysical inconsistency in addition to logical inconsistency. For although a maximal, *metaphysically* inconsistent world need not verify a C-, CF-, or NCF-inconsistency, it remains true that any world that verifies the antecedent in (5) *does*. This is all that is needed for Bjerring's argument to go through.

Partial Modal Space: Impossible worlds correspond to arbitrary (maximal or partial) inconsistent sets of sentences.

Stratified Modal Space: Impossible worlds correspond to (maximal or partial) sets of sentences that are closed and consistent with respect to some non-classical logic.

We can think of stratified modal space as a collection of subspaces, each subspace containing all and only worlds that are governed by the logic specific to that subspace. In particular, the set of possible worlds corresponds to the subspace containing all and only worlds that are governed by classical logic (assuming that classical logic is actually true). By contrast, partial modal space has no internal structure: it is construed as a single space containing all inconsistent sets of sentences in addition to the set of possible worlds.

Which of these alternative world-ontologies should we opt for? Bjerring argues that partial modal space is too "permissive" or "unconstrained" as it makes room for a kind of misbehaved worlds that fail to "reflect the nontrivial [...] logical dependencies or relations that obtain between various sentences" (Bjerring 2014, 344). For instance, nothing in the construction of partial modal space excludes an impossible world in which "intuitionistic logic and $A \wedge B$ are true, but A or B not" (Bjerring 2014, 344). Such a world fails to reflect the fact that conjunction elimination is valid in intuitionistic logic. In turn, the objection goes, it is not clear how partial modal space is geared to capture the kinds of non-trivial principles that govern our reasoning about impossibilities. By contrast, stratified modal space is well-suited for this job, since it rules out misbehaved worlds from the get-go by requiring that any world that verifies a logic $\mathcal{L}$ is closed and consistent with respect to $\mathcal{L}$.

Let us concede that misbehaved worlds should be excluded from extended modal space.[11] Does this give us reason to reject partial modal space as Bjerring argues? No, since Bjerring's objection conflates two conceptions of what it means for a logic to be true at a world:

---

[11] In fact, one need not grant this point: misbehaved worlds arguably cause no harm as long as our closeness metric on extended modal space ensures that we rarely (or perhaps never) have to consult misbehaved worlds to evaluate counterfactuals. To illustrate this point, suppose w is a "behaved" world that verifies {"paraconsistent logic is true", "the principle of explosion is invalid"}, whereas w' is a "misbehaved" world that verifies {"paraconsistent logic is true", "the principle of explosion is valid"}. Consider then a world w_u in which (4) is uttered. If our closeness metric ensures that w is closer to w_u than w', nothing prevents ELS from delivering the correct semantic verdict for (4) even if w' is contained in extended modal space. So it is far from clear that the mere inclusion of misbehaved worlds in extended modal space is problematic.

$\mathcal{L}$-**membership**: A logic $\mathcal{L}$ is true at a world w if and only if the sentence "$\mathcal{L}$ is true" is a member of w.

$\mathcal{L}$-**closure**: A logic $\mathcal{L}$ is true at a world w if and only if w is closed and consistent with respect to $\mathcal{L}$.

$\mathcal{L}$-membership and $\mathcal{L}$-closure are non-equivalent conceptions of what it means for a logic to be true at a world. To see this, consider a world w_{IL} that contains the three sentences "intuitionistic logic is true", $\varphi \wedge \psi$, and $\neg\varphi$. Is intuitionistic logic true in w_{IL}? The answer depends on whether we regard $\mathcal{L}$-membership or $\mathcal{L}$-closure as the correct account of what it means for a logic to be true at a world. If $\mathcal{L}$-membership is correct, the answer is "yes" since "intuitionistic logic is true" is a member of w_{IL}. If $\mathcal{L}$-closure is correct, the answer is "no" since w_{IL} is not closed and consistent with respect to intuitionistic logic. So $\mathcal{L}$-membership and $\mathcal{L}$-closure are not equivalent, and thus cannot both be used to account for what it means for a logic to be true at a world.

To see why Bjerring's objection against partial modal space conflates $\mathcal{L}$-membership and $\mathcal{L}$-closure, notice, as a first step, that stratified modal space rules out misbehaved worlds akin to w_{IL} only if $\mathcal{L}$-closure is correct. For if $\mathcal{L}$-membership were correct, nothing would prevent w_{IL} from being part of stratified modal space. The reason is that a set of sentences may well have "intuitionistic logic is true" as a member while failing to be closed and consistent with respect to intuitionistic logic. Conversely, a set of sentences may well have "intuitionistic logic is false" as a member while being closed and consistent with respect to intuitionistic logic. There is no necessary connection between the particular sentences included in a set and the logical structure of the set as a whole. So if $\mathcal{L}$-membership were correct, stratified modal space would not rule out misbehaved worlds. In turn, Bjerring is committed to $\mathcal{L}$-closure if he wants to maintain that stratified modal space *does* rule out such misbehaved worlds.

A commitment to $\mathcal{L}$-closure should not, by itself, worry Bjerring. In fact, Bjerring at one point articulates a view that sits nicely with $\mathcal{L}$-closure:

> [I]f we are asked to evaluate what would have happened, had some non-classical logic $\mathcal{L}_i$ been correct, we are asked to consider what happens in a world whose truths and logical features cannot adequately be described or codified by classical logic. We are not merely asked to consider what happens in a world that verifies sentences such as "$\mathcal{L}_i$ is the correct logic" rather than "classical logic is the correct logic," but whose logical structure otherwise is identical to the one in our world. Rather, we consider what happens in a logically impossible world whose truths are governed by $\mathcal{L}_i$ rather than classical logic (Bjerring 2014, 334).

I am inclined to agree with Bjerring that $\mathcal{L}$-closure is preferable to $\mathcal{L}$-membership, although I will not pause to defend the view here. The point I wish to make here is that

Bjerring's worry about partial modal space can be dismissed if $\mathcal{L}$-closure is correct. For if $\mathcal{L}$-closure is correct, there simply *are* no misbehaved worlds for partial modal space to include. The (somewhat trivial) reason is that a set of sentences cannot both *be* and *not be* closed and consistent with respect to a given logic. In turn, if $\mathcal{L}$-closure is correct, no world can verify a logic $\mathcal{L}$ and fail to be closed and consistent with respect to $\mathcal{L}$. In particular, no world can verify intuitionistic logic and $\varphi \wedge \psi$ while failing to verify φ or ψ. So if $\mathcal{L}$-closure is true, the kind of misbehaving worlds that Bjerring is worried about do not exist, and, *a fortiori*, will not be part of partial modal space. The upshot is that Bjerring's commitment to $\mathcal{L}$-closure means that his objection against partial modal space can be dismissed. In effect, this leaves the choice between partial modal space and stratified modal space entirely open.

3.3 In defense of partial modal space

I will now tip the balance in favour of partial modal space by arguing that stratified modal space imposes too strict constraints on what counts as an impossible world. The first thing to note is that partial modal space is maximally unconstrained as it validates (a slightly generalized version of) Nolan's "comprehension principle" (Nolan 1997: 542):

> **Comprehension Principle (CP)**: For any set of propositions Γ that cannot be jointly true, there is an impossible world where all and only the sentences in Γ are true.

Intuitively, CP expresses an "anything goes" view of impossible worlds: there are no constraints on how logically ill-behaved impossible worlds are allowed to be. Most accounts of impossible worlds invalidate CP as they impose some sort of non-trivial closure constraint on impossible worlds. Such accounts may roughly be divided into two categories. On the one hand, we have accounts that weaken the classical consequence relation in a *uniform* manner: impossible worlds are required to obey the laws of a single non-classical logic.[12] On the other hand, we have accounts—akin to stratified modal space—that weaken the classical consequence relation in a *non-uniform* manner: impossible worlds are required to obey the laws of *some* non-classical logic, but different impossible worlds may obey the laws of different non-classical logics. I will now argue that both "uniform" and "non-uniform" accounts of impossible worlds are inadequate for the purposes of developing a general semantics of all counterfactuals.

---

12 See Mares (1997) and Restall (1997) examples of such an account of impossible worlds.

The reason why "uniform" accounts of impossible worlds fail is that any uniform weakening of the classical consequence relation will result in a modal space that excludes at least *some* impossible scenarios that we may be able to reason non-trivially about.[13] To illustrate this point, consider a non-classical logic $\mathcal{L}_\neg$ that Fagin et al. (1995, 321–325) discuss in an epistemic logic context. $\mathcal{L}_\neg$ differs from classical logic by replacing the classical truth-functional semantics of negation ($\neg\varphi$ is true if and only if φ is false) with a non-truth-functional semantics that assigns independent truth-values to $\neg\varphi$ and φ.[14] Fagin et al. show that this non-classical semantics of negation implies that $\{\varphi \rightarrow \psi, \varphi\}$ no longer entails. By contrast, $\{\varphi, \psi\}$ still entails $\varphi \wedge \psi$. Now, suppose we were to close impossible worlds under logical consequence in $\mathcal{L}_\neg$. Consider, then, the following conditional:

(7)   If all valid inferences of $\mathcal{L}_\neg$ were invalid, conjunction introduction would be valid.

(7) is non-trivially false since, if all valid inferences of $\mathcal{L}_\neg$ were invalid, then, in particular, conjunction introduction would be invalid. However, if impossible worlds are required to obey the laws of $\mathcal{L}_\neg$, no impossible world will verify the antecedent of (7). In turn, ELS will fail to deem (7) non-trivially false. So it turns out to be a bad idea to require impossible worlds to obey the laws of $\mathcal{L}_\neg$.

To be sure, the point generalizes in the straightforward way: let $\mathcal{L}$ be any preferred non-classical logic, and suppose impossible worlds are required to obey the laws of $\mathcal{L}$. ELS will then fail to deem the following counterpossible nontrivially false:

(8)   If all valid inferences of $\mathcal{L}$ were invalid, then some valid inference of $\mathcal{L}$ would be valid.

So if impossible worlds are required to obey the laws of a single non-classical logic, extended modal space will end up being too constrained.[15]

Essentially the same objection can be raised against "non-uniform" accounts of impossible worlds (akin to stratified modal space). Let $\mathcal{L}_1, \ldots, \mathcal{L}_n$ be a set of logics such that every impossible world is required to obey the laws of

---

13 Nolan (1997, 546–547) also makes a point along these lines.
14 One natural way of implementing this idea formally is by replacing the classical truth assignment which assigns independent truth-values to all and only *atomic* sentences with a non-classical truth assignment which assigns independent truth-values to all and only *literals*, where a literal is an atomic sentence or its negation.
15 Brogaard and Salerno (2013, 651) make a very similar point.

$\mathcal{L}_i$, for some i. Furthermore, let $|\mathcal{L}_i|$ denote the set of valid inferences in $\mathcal{L}_i$. Consider, then, the following conditional:

(9)   If all inferences in $|\mathcal{L}_1| \cup \cdots \cup |\mathcal{L}_1|$ were invalid, then some inference in $|\mathcal{L}_1| \cap \cdots \cap |\mathcal{L}_1|$ would be valid.

(9) is non-trivially false since, if all inferences in $|\mathcal{L}_1| \cup \cdots \cup |\mathcal{L}_1|$ are invalid, it follows, *a fortiori*, that all inferences in $|\mathcal{L}_1| \cap \cdots \cap |\mathcal{L}_1|$ are invalid. But if all impossible worlds must obey the laws of $\mathcal{L}_i$, for some i, it follows that no impossible world will verify the antecedent in (9). As such, ELS will fail to deem (9) non-vacuously false. So Bjerring's stratified modal space turns out to be too constrained.

Bjerring (2014, 350) anticipates this kind of objection by pointing out that nothing prevents us from including in stratified modal space a completely trivial logic $\mathcal{L}_{triv}$ with no principles governing logical consequence. This is essentially a way of ensuring that stratified modal space validates CP. In turn, some impossible world will verify the antecedent in (9); and since any world that verifies the antecedent in (9) will falsify its consequent, ELS will rightly deem (9) non-trivially false.

The inclusion of $\mathcal{L}_{triv}$ means that stratified modal space becomes extensionally equivalent to partial modal space (both modal spaces validate CP). Why, then, should we prefer stratified modal space over partial modal space? Bjerring (2014: 344-45) seems to cite the internal structure of stratified modal space as a deciding factor: by grouping impossible worlds together according to which logics they obey, we get an interesting structure on extended modal space that reflects the various principles that govern our reasoning about impossibilities.

I think this line of reasoning should be resisted: it is neither necessary nor desirable to impose an internal structure on extended modal space. Let me explain why by drawing a parallel to the standard possible worlds framework. Standard modal space is usually construed as a single unstructured set of worlds (e.g. the set of all maximal, consistent sets of sentences in some appropriate language). In principle, we *could* impose an internal structure on standard modal space, e.g. by grouping possible worlds together according to which laws of physics they obey. But there is no need to build this kind of structure into the construction of standard modal space since our *closeness metric* provides all the structure that is needed for LS to deliver its semantic verdicts. As long as our closeness metric on standard modal space makes proper mentioning of various laws of nature, there is no theoretical gain in imposing a structure on standard modal space. Similarly, as long as our closeness metric on extended modal space makes proper mentioning of various laws of logic, there is no theoretical gain in imposing a structure on extended modal space.

In reply, Bjerring might argue that it becomes easier to formulate closeness principles if we can appeal to a structure within extended modal space. Consider, e.g. a principle which Bjerring suggests should govern relative closeness among worlds in different subspaces of stratified modal space:

> (**Relative Closeness Condition**) For any counterfactual whose antecedent presupposes that some logic $\mathcal{L}_i$ is correct (true, adequate), a world in modal space $W_{\mathcal{L}_i}$ [the set of worlds that obey the laws of $\mathcal{L}_i$] is closer to the actual world than any world in modal space $W_{\mathcal{L}_j}$ where $W_{\mathcal{L}_i} \neq W_{\mathcal{L}_j}$ (Bjerring 2014, 348).

The intuitive idea underlying (Relative Closeness Condition) is that "if some logic $\mathcal{L}_i$ had indeed been correct, then regardless of what else might have been the case, the laws of $\mathcal{L}_i$ would have been the case" (Bjerring 2014, 348). Suppose we accept this closeness principle. The question is whether we can state (Relative Closeness Condition) without appeal to the internal structure of stratified modal space. It seems so:

> **Relative Closeness Condition***: For any counterfactual whose antecedent presupposes that a logic $\mathcal{L}_i$ is true, any world that obeys $\mathcal{L}_i$ is closer to the world of utterance than any world that does not.

Relative Closeness Condition* is equivalent to (Relative Closeness Condition), and it is hard to see why (Relative Closeness Condition) should be preferable to Relative Closeness Condition*. If anything, Relative Closeness Condition* is simpler than (Relative Closeness Condition). So it does not look like the internal structure of stratified modal space helps to simplify our closeness metric.

Finally, note that a closeness metric should not only capture *logical* reasoning, but also, e.g. *metaphysical* and *mathematical* reasoning. In these other areas, our closeness metric cannot appeal to the internal structure of stratified modal space (unless we built additional structure into the construction of stratified modal space). In turn, there is a risk that our closeness metric will end up being something of a heterogeneous construct that sometimes, but not always, appeals to the structure of extended modal space. A more promising strategy, it seems to me, is to avoid imposing any internal structure on extended modal space—which amounts to adopting partial modal space—and let our closeness metric deliver the required structure.

4 Concluding remarks

The extended Lewis-Stalnaker semantics promises to offer a non-vacuous account of counterpossibles by quantifying over both possible and impossible

worlds. Linguistic ersatz theorists often construe impossible worlds as maximal, inconsistent sets of sentences in some appropriate language. However, Bjerring has shown that the extended Lewis-Stalnaker semantics delivers the wrong semantic verdicts for many counterpossibles if impossible worlds are required to be maximal. To make room for partial impossible worlds, Bjerring considers two alternative world-ontologies: partial modal space and stratified modal space. In partial modal space, impossible worlds correspond to arbitrary (maximal or partial) inconsistent sets of sentences, whereas, in stratified modal space, they correspond to (maximal or partial) sets of sentences that are closed and consistent with respect to some non-classical logic. Bjerring argues that partial modal space is too permissive or unconstrained. I have argued that this objection conflates two different conceptions of what it means for a logic to be true in a world. Second, I have argued that stratified modal space imposes too strong constraints on what counts as an impossible world. In sum, linguistic ersatzists should adopt partial modal space and construe impossible worlds as arbitrary inconsistent sets of sentences.

References

Adams, R. (1974). "Theories of Actuality". *Nous* 8: 211–231.

Bjerring, J.C. (2014). "On Counterpossibles". *Philosophical Studies* 4: 327–353.

Brogaard, B. and J. Salerno (2013). "Remarks on Counterpossibles". *Synthese* 190(4): 639–660.

Fagin, R. et al. (1995). *Reasoning About Knowledge*. Cambridge: MIT Press.

Goodman, J. (2004). "An Extended Lewis-Stalnaker Semantics and The New Problem of Counterpossibles". *Philosophical Papers* 33(1): 35–66.

Jago, M. (2013). "Impossible Worlds". *Noûs* 47(3): 713–728.

Jago, M. (2014). *The Impossible: An Essay on Hyperintensionality*. Oxford: Oxford University Press UK.

Laan, D. Vander (1997). "The Ontology of Impossible Worlds". *Notre Dame Journal of Formal Logic* 38(4): 597–620.

Laan, D. Vander (2004). "Counterpossibles and Similarities". In: F. Jackson and G. Priest (eds.) *Lewisian Themes: The Philosophy of David K. Lewis*. Oxford: Clarendon Press.

Lewis, D. (1973). *Counterfactuals 2*. Oxford: Blackwell Publishers, pp. 278–281.

Mares, E. (1997). "Who's Afraid of Impossible Worlds?" *Notre Dame Journal of Formal Logic* 38(4): 516–526.

Nolan, D. (1997). "Impossible Worlds: A Modest Approach". *Notre Dame Journal of Formal Logic* 38(4): 535–572.

Nolan, D. (2013). "Impossible Worlds". *Philosophy Compass* 8(4): 360–372.

Restall, G. (1997). "Ways Things Can't Be". *Notre Dame Journal of Formal Logic* 38(4): 583–596.

Stalnaker, R. (1968). "A Theory of Conditionals". *In Studies in Logical Theory, American Philosophical Quarterly*, Monograph Series, 2. Oxford: Blackwell, 98–112.