Abstract

Fictionalists maintain that possible worlds, numbers or composite objects exist only according to theories which are useful but false. Hale, Divers and Woodward have provided arguments which threaten to show that fictionalists must be prepared to regard the theories in question as contingently, rather than necessarily, false. If warranted, this conclusion would significantly limit the appeal of the fictionalist strategy rendering it unavailable to anyone antecedently convinced that mathematics and metaphysics concern non-contingent matters. I try to show that their arguments can be resisted by developing and defending a strategy suggested by Rosen, Nolan and Dorr, according to which the fiction-operator is to be analysed in terms of a counterfactual that admits of non-trivial truth-values even when the antecedent is impossible.

1 Fictionalism and the Contingent Fiction Argument

Fictionalists about worlds (following Rosen 1990), numbers (following Field 1980, 1989) or mereological sums (following Dorr & Rosen 2002) consider theories such as modal realism, standard arithmetic or compositional universalism to be useful but false. Contingently false or necessarily false? On the face of it, fictionalists had better be able to regard these theories as necessarily false. After all they contain, respectively, postulates of modal metaphysics, mathematical axioms and composition principles, which count among the best candidates for statements that have their truth-value as a matter of metaphysical necessity. The Contingent Fiction Argument (CFA), however, threatens to show that an important class of fictionalists are not in fact entitled to regard their fiction as impossible. If successful, the CFA would significantly limit the appeal of these fictionalist
strategies: they would no longer be an option for those who think that mathematics
and metaphysics concern non-contingent matters.

While early variants of the CFA are due to Hale (1995: 65) and Divers (1999:
335), the most detailed and general version is provided by Woodward (2010). On
this version, the CFA is directed at fictionalist positions that exhibit the following
three features. First, the fictionalists take their fiction to provide a standard of
correctness for sentences belonging to the target discourse. For instance, modal
fictionalists account for the fact that ‘there is a world containing a blue swan’ is
deemed correct, while ‘there is a world containing a round square’ is not, by tying
correctness to truth-in-the-relevant-possible-world-fiction. Mathematical and
compositional fictionalists account for the (in)correctness of sentences quantifying
over numbers and composites in an analogous way. Accordingly, Woodward
(2010: 411) takes the fictionalists to endorse (the instances of) the following
schema:

(F) ‘S’ is correct ⇔ According to F, [S]T

Here ‘F’ stands for whichever theory is regarded as fiction, e.g. Lewis’s modal
realism, and ‘[S]T’ is the regimentation of the target sentence ‘S’ in the canonical
language of that theory, e.g. the language of counterpart theory.

Second, the fictionalists follow Rosen (1990: §8) in analyzing the fiction-operator
in terms of a counterfactual conditional (where ‘explicit F’ is a sentence stating
the explicit content of F):


Third, the fictionalists accept the following standard principle for counterfactuals
(see e.g. Lewis 1973: 24-26):

(Trivialisations) Any counterfactual with an impossible antecedent is true.

In its simplest form, the CFA then consists in the observation that fictionalists who
accept (F), (Operator) and (Trivialisation) cannot regard their fiction as impossible.
If they did, they would unacceptably be forced to regard any sentence whatsoever
of the target discourse as correct.

2In discussions of modal fictionalism, the schema is often simply put as follows:

(F*) S ⇔ According to F, [S]T

As should become clear, employing (F*) instead of (F) does not help with respect to the CFA. A
reason to prefer (F) over (F*) in the present context is that (F*) appears to force fictionalists to
accept unqualified existence claims concerning worlds/numbers/composites as straightforwardly
true, which in turn appears to force them to regard those claims as expressing meta-fictional
propositions (see Liggins 2008), a commitment some of our fictionalists might prefer to avoid.
2 The strengthened Contingent Fiction Argument

A promising response to the simple CFA is the *counterpossible strategy*: give up (Trivialisation) by allowing for non-trivial counterfactuals with impossible antecedents (see Rosen 1995: 69-70, Nolan 2005: 3, Dorr 2008: §2). The CFA, however, can be strengthened: following Divers (1999), Woodward (2010) argues that the counterpossible strategy conflicts with other fictionalist desiderata.

Consider modal fictionalists’ aspiration to perform ‘modal logic by proxy’ (Divers 1999: §V): when confronted with an argument couched in modal terms, they will want to consider the corresponding argument couched in counterpart-theoretic terms and will want to take the validity of the latter to reveal the validity of the former. Fictionalists can go some way towards justifying this practice by pointing out that they endorse not only worldly instances of (F), i.e. instances that result from replacing ‘S’ with a sentence explicitly quantifying over possible worlds, but also modal instances, i.e. instances that result from replacing ‘S’ with a modal sentence. By themselves, however, modal instances of (F) do not provide a guarantee that whenever there is counterpart-theoretic entailment, then there also is an entailment among the corresponding modal sentences.

Similarly, consider fictionalists’ unqualified use of sentences explicitly quantifying over the entities they reject. Suppose a given counterpart-theoretic sentence, ‘[S₁]ᵣ’, holds according to the fiction. By a worldly instance of (F), the corresponding possible-world sentence ‘S₁’ is correct. Now, suppose that a different counterpart-theoretic sentence, ‘[S₂]ᵣ’, follows from ‘[S₁]ᵣ’. Fictionalists want a guarantee that the possible-world sentence ‘S₂’ which corresponds to ‘[S₂]ᵣ’ will be correct too. But, by themselves, worldly instances of (F) don’t provide one. The point clearly generalizes to mathematical and compositional fictionalism: fictionalists need a guarantee that entailments are exportable from the fiction.

Following Divers, Woodward argues that fictionalists can provide this guarantee if they can establish a *Safety Result*:

(Safety) If ‘[S₁]ᵣ’ entails ‘[S₂]ᵣ’ then it follows that the correctness of ‘S₁’ entails the correctness of ‘S₂’.

The strengthened CFA then maintains that (Safety) can be established only on assumptions that conflict with the counterpossible strategy. To see this, consider how Woodward derives (Safety). First, he gives a modal account of the notion of entailment featuring in (Safety), so that it amounts to:

(Safety*) □(□([S₁]ᵣ → [S₂]ᵣ) → □(‘S₁’ is correct → ‘S₂’ is correct))

Crucially, he then assumes counterfactuals to be closed under modal entailment:
(Mod-Closure) If \( A \rightarrow B \) and \( \Box(B \rightarrow C) \), then \( A \rightarrow C \)

Given these assumptions, (Safety*) can be established as follows. First we make two further assumptions later to be discharged:

1. \( \Box([S_1]_T \rightarrow [S_2]_T) \)
2. ‘\( S_1 \)’ is correct

From (2) and the relevant instances of (F) and (Operator) we derive:

3. \( \text{Explicit}_F \rightarrow [S_1]_T \)

Given (1), (Mod-Closure) yields:

4. \( \text{Explicit}_F \rightarrow [S_2]_T \)

Working backwards, we appeal to (Operator) and (F) arriving at:

5. ‘\( S_2 \)’ is correct.

We discharge (2) to derive:

6. ‘\( S_1 \)’ is correct \( \rightarrow \) ‘\( S_2 \)’ is correct.

To arrive at the desired necessitation of this conditional we note that all the assumptions relied upon to derive (6) plausibly hold with necessity (provided they hold at all), so that we could have equally assumed their necessitations. We then rely on the principle, which holds in S4, that if \( P \) follows from \( Q_1, ..., Q_n \) each of which has \( \Box \) as its main operator, then so does \( \Box P \). We thus infer:

7. \( \Box(‘S_1’ \rightarrow ‘S_2’) \)

Discharging (1) and applying the same reasoning that led from (6) to (7) we derive:

(Safety*) \( \Box(\Box([S_1]_T \rightarrow [S_2]_T) \rightarrow \Box(‘S_1’ \rightarrow ‘S_2’) \)

What conflicts with the counterpossible strategy is (Mod-Closure). For it can hardly be doubted that the fictionalists’ fiction counterfactually implies itself: \( \text{Explicit}_F \rightarrow \text{Explicit}_F \). And given that the fiction is impossible, ‘\( \text{Explicit}_F \)’ strictly implies any sentence whatsoever. As per (Trivialisation), it then follows from (Mod-Closure) that ‘\( \text{Explicit}_F \)’ counterfactually implies any sentence whatsoever and the CFA is back on track.
3 Rejecting the strengthened Contingent Fiction Argument

Can fictionalists make do without (Mod-Closure)? The point of the Safety Result, remember, is to ensure that the fictionalists’ standard of correctness accurately reflects the entailment relations among sentences belonging to counterpart theory, compositional universalism or standard arithmetic. Now, the theories in question are typically taken to be governed by a relation of *logical* entailment, rather than by mere strict implication. For instance, one of the main reasons why modal fictionalists want to perform modal logic by proxy is precisely that this way the validity of modal arguments reduces to first-order logical entailment among counterpart-theoretic sentences (see Divers 1999: 329; Woodward 2010: 410). Similar points apply to mereological and mathematical theories, which are typically taken to consist of mereological and mathematical axioms and all their *logical* consequences. It is therefore dialectically permissible for fictionalists to focus on such logical entailments and, accordingly, to replace (1) with:

\[(1^*) \left[ S_1 \right]_T \vdash \left[ S_2 \right]_T\]

All that is required to derive the correspondingly modified Safety Result is then closure under *logical* entailment:

**(Log-Closure)** If \( A \Box \to B \) and \( B \vdash C \), then \( A \Box \to C \)

By retreating to (Log-Closure) fictionalists gain the right to regard their fiction as *metaphysically* impossible. Of course they still have to regard it as *logically* possible. But this they will do anyway. Modal fictionalists, for instance, would hardly want to study the validity of modal arguments by way of considering counterpart-theoretic proxies if they regarded counterpart theory as inconsistent. While the present refinement of the counterpossible strategy thus seems promising, it also faces two problems (Woodward 2010: 415-6).

The **Content Problem** arises because modal fictionalists need certain ‘microreduction laws’, i.e. metaphysically necessary truths connecting micro- and macro-reality, to hold according to their possible-world-fiction PW. For example, let ‘Micro’ describe the microphysical structure of a world containing a blue swan. PW’s explicit content - in the form of a recombination principle operating on spacetime points and fundamental properties - ensures that, according to PW, there is a world at which ‘Micro’ is true. But, on its own, this doesn’t yet ensure that it is true according to PW that there is a world which contains a blue swan. Rather, it must additionally be ensured that the microreduction law ‘Micro \( \to \) there is blue swan’ holds according to PW. While this is guaranteed by (Mod-Closure) it is not clear that (Log-Closure) will do too.
The Parity Problem points to a conflict between a simultaneous rejection of (Mod-Closure) and retention of (Log-Closure). To reject (Mod-Closure) is to allow at least some counterfactuals with metaphysically impossible antecedents to receive non-trivial truth-values. To accept (Log-Closure) is to maintain that all counterfactuals with logically impossible antecedents are still trivially true. But, according to Woodward, whoever takes seriously the possibility of non-trivial countermetaphysical reasoning should also take seriously the possibility of non-trivial counterlogical reasoning. The present strategy, so the accusation, flouts this principle of parity.

The key to addressing both problems is to appreciate that a phenomenon well known from ‘ordinary’ counterfactuals carries over to counterpossibles. Thus, when contemplating what Nixon’s button pressing would have resulted in, we famously hold onto the actual laws of nature as much as feasible (see Lewis 1979: 467-472) and accept:

(8) If Nixon had pressed the button, there would have been a nuclear holocaust.

Now, something very similar applies to our evaluation of counterpossibles. To suppose that Hobbes succeeded in squaring the circle is to suppose that some actual laws of mathematics are violated. Radical though such a violation may be, we still hold on to as many actual laws as we can when reasoning under that supposition. As Nolan points out (1997: 544), we tend to reject:

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

A plausible explanation of this is that the truth of the consequent would require not only a violation of the laws of mathematics but an additional violation of the laws of logic. Since we avoid such additional violations when considering what the truth of the antecedent would have resulted in, we regard the counterfactual as false.

Note that an additional violation of laws doesn’t have to be a violation of an additional type of law. While the above example is of this form (violation of mathematical laws vs. violation of mathematical and logical laws) we also get cases where we have more and less severe violation of laws of one and the same type. Thus Nolan points out that we have substantial knowledge about what the laws of logic would be ‘if intuitionistic logic were the One True Logic’ (1997: 545). For instance, we accept

(10) If intuitionism were correct, the law of excluded middle wouldn’t be valid, while rejecting
(11) If intuitionism were correct, the law of non-contradiction wouldn’t be valid.

The reason is that the truth of the consequent in (11) does, while the truth of the consequent in (10) does not, require a gratuitous additional deviation of the laws of logic not required by the truth of the antecedent.

Fictionalists can now exploit this phenomenon to solve both the content and the parity problem. Starting with the former, let ‘Law’ express a given microreduction law. The challenge for modal fictionalists then is to justify:

**(Micro-Macro)** According to PW, Law.

Woodward is right that if fictionalists want to support (Micro-Macro) by relying on a closure principle then it’s got to be (Mod-Closure) rather than (Log-Closure). For in that case fictionalists exploit the necessity of microreduction laws which ensures:

(12) □(ExplicitPW → Law)

Then they rely on the fiction counterfactually implying itself:

(13) ExplicitPW □→ ExplicitPW

In combination with (Mod-Closure) it then follows:

(14) ExplicitPW □→ Law

From this they derive (Micro-Macro) by the relevant instance of (Operator). This reasoning can indeed not be replicated with (Log-Closure) in place of (Mod-Closure). For the assumption which would then have to replace (12) is implausible: microreduction laws hold of metaphysical necessity, not as a matter of logic, and cannot be derived from PW with the help of logic alone.

But why think that the fictionalists’ justification of (Micro-Macro) has to take this form? Counterfactuals can be true even though the antecedent doesn’t entail the consequent, modally or otherwise. So it should be possible to address the content problem by providing a more direct reason for regarding (14) as true, one that does not involve a detour via an entailment claim and a matching closure principle. Fortunately, the role which laws play for our counterfactual thinking in general constitutes just such a reason: that PW is metaphysically impossible doesn’t mean that no metaphysical laws at all would survive its truth, just like it would be wrong to think that no logical laws at all would survive the truth of intuitionism. On the contrary, since there is no tension between modal realism and the microreduction laws, the default assumption should be that the microreduction laws would still
have held, had \( PW \) been true, just like we think that the law of non-contradiction would still have held, had intuitionism been correct.

Turning to the parity problem, first note that (Log-Closure) is still stronger than it needs to be for the fictionalists’ purposes. For (Log-Closure) says that any counterfactual you like is closed under logical entailment. To establish the Safety Result, however, fictionalists only need a special class of counterfactuals to be closed under logical entailment: those whose antecedents state the explicit content of one of the fictions in question. Thus fictionalists can further retreat to the following weakening of (Log-Closure):

\[(W\text{-Log-Closure}) \quad \text{If } \text{Explicit}_F \rightarrow B \text{ and } B \vdash C \text{, then } \text{Explicit}_F \rightarrow C\]

In line with the parity principle, fictionalists can then accept that the stronger (Log-Closure) doesn’t hold any more than (Mod-Closure) and, accordingly, that non-trivial counterlogical reasoning is just as possible as non-trivial countermetaphysical reasoning. Of course, it is now fair to ask our fictionalists to provide a reason why counterpossibles with the relevant fictions as antecedents should be closed under logical entailment, given that counterpossibles in general are not. But, again, the role which laws play for counterfactual thinking constitutes precisely such a reason. For recall that, according to our fictionalists, the relevant fictions - modal realism, standard arithmetic, compositional universalism - while metaphysically impossible are perfectly consistent with the actual laws of logic. In line with the general rule that we avoid gratuitous violations of laws when reasoning counterfactually (and continue to do so in the special case when we reason counterpossibly), our fictionalists are thus justified in maintaining that if their preferred fiction had been true, the laws of logic would have been the same as they actually are. So given that, in fact, the truth of ‘\( B \)’ logically ensures the truth of ‘\( C \)’, it would still have done so had the fiction been true. Given further that ‘\( B \)’ would have been true if the fiction had been true, the fictionalists are entitled to conclude that ‘\( C \)’ would have been true too had the fiction been true.

This response doesn’t burden our fictionalists with assumptions that they, qua proponents of non-trivial counterpossibles, aren’t likely to make anyway. Thus consider Nolan’s \textit{Strangeness of Impossibility Condition} which enjoys wide popularity on accounts of non-trivial counterpossibles:

\[(SIC) \quad \text{For any possible world } w, \text{ any possible world is more similar (nearer) to } w \text{ than any impossible world}^{[3]}\]

\(^{[3]}\)In the informal discussion Nolan focusses on a corollary of (SIC) (1997: 550); in the formal appendix it becomes apparent that he is really concerned with (SIC) (566).
Nolan motivates (SIC) as being based on the intuitively plausible idea that ‘the heavens will fall before (correct) logic will fail us’ (550). From Nolan’s surrounding discussion it is, however, clear that the ‘impossible worlds’ in (SIC) cover logically and merely metaphysically impossible worlds. Thus (SIC) really reflects the idea that the heavens will fall before the principles of either metaphysics or logic will fail us. What corresponds more directly to Nolan’s motivation is a Strangeness of Logical Impossibility Condition:

(SLIC) For any possible world \( w \), any logically possible world is more similar (nearer) to \( w \) than any logically impossible world.\(^4\)

Anyone sympathetic to (SIC) should also be sympathetic to (SLIC): in tandem the conditions reflect the plausible idea that the heavens will fall before principles of metaphysics will fail us, which in turn will happen before logic will fail us. And (SLIC) guarantees (W-Log-Closure). For it ensures that the logically possible worlds form a sphere around the actual world. The truth of ‘\( B \vdash C \)’ then ensures that ‘\( B \rightarrow C \)’ holds throughout that sphere. Given that ‘\( \text{Explicit}_F \)’ is logically possible, the closest \( \text{Explicit}_F \)-worlds are guaranteed to be within the sphere. The truth of ‘\( \Box \text{Explicit}_F \rightarrow B \)’ ensures that these closest \( \text{Explicit}_F \)-worlds are \( B \)-worlds. It then follows that the closest \( \text{Explicit}_F \)-worlds are also \( C \)-worlds, which means that ‘\( \text{Explicit}_F \rightarrow C \)’ is true as well.\(^5\)

A final issue needs to be addressed before fictionalists have earned the right to rely on the Safety Result along the present lines. For recall that at later stages in the derivation (e.g. the move from (6) to (7)) we needed to assume that all principles appealed to hold necessarily. So fictionalists in fact need instances of (W-Log-Closure) not only to express a truth, but a necessity. But, again, this doesn’t go beyond what proponents of the counterpossible strategy will likely accept anyway. In fact, (SLIC) ensures this as well, since it ensures not only that the actual world is surrounded by a sphere of all logically possible worlds, but ensures that any possible world is.

\(^4\)(SIC) and (SLIC) are conditions on similarity orderings of worlds proposed as part of a worlds-based semantics for counterfactuals, a semantics with which fictionalists about non-actual worlds are likely to be unhappy. This, however, is not a problem for the present solution. What matters for that solution is that all of our fictionalists can rely on (W-Log-Closure) which doesn’t presuppose the adequacy of a worlds-based semantics for counterfactuals. The present discussion of the worlds-based conditions (SIC) and (SLIC) is meant to support the claim that (W-Log-Closure) is not a hefty additional assumption in the present context by showing that theorists who are sympathetic to non-trivial counterpossibles in the first place - theorists which happen not to be fictionalists about worlds - are also sympathetic to assumptions which guarantee (W-Log-Closure).

\(^5\)For ease of presentation it is assumed that the truth-conditions for a counterfactual ‘\( A \rightarrow B \)’ can be formulated in terms of the closest \( A \)-worlds being \( B \)-worlds, a formulation which relies on the Limit Assumption (see Lewis 1973: 19-20). This assumption is not essential to the argument.
I will close by addressing a potential problem for the proposed solution. We noted that (W-Log-Closure) can be based on (SLIC) which naturally supplements the widely held (SIC). But while (SIC) is often endorsed by its proponents without qualification (Berto et al. 2017: §§2.3, 3.1; Jago 2015: 714; Kment 2006a: 301; Mares 1997: 521), it is sometimes put forward rather as a default principle that holds ordinarily but may break down in extra-ordinary contexts (Kment 2006b: §§2.3-2.4, Nolan 1997: 550-1, Van der Laan 2004: 271-2). What are the reasons for this more cautious stance on (SIC) and do they pose a problem for the present strategy?

When (SIC) is proposed as a mere default principle this is typically done due to the worry that there might be specific counterexamples: counterfactuals with a possible antecedent and an impossible consequent which, in the right context, may still sound true (or at least not clearly false). One of the less contrived cases is Nolan’s (551):

(15) If intuitionistic logic came to be thought a much more satisfactory basis for mathematics by the experts, and if intuitionistic investigations led to breakthroughs in many areas of inquiry [...], then intuitionistic logic would turn out to be correct after all.

Confronted with such cases, there are at least two promising lines of response available to fictionalists who want to preserve the right to rely on (W-Log-Closure), and hence on the Safety Result, without qualification.

First, it is far from clear that the linguistic data is sufficiently robust to warrant tinkering with (SIC) and (SLIC). In the case at hand, Nolan admits that there might be no conflict after all, since we might only be judging (15) as correct insofar as we take it to express an epistemic modality (roughly: as saying something about what our evidence would have let us to conclude if it had been different in certain respects, rather than what the world would have been like if it had been different in certain respects). A full defence along these lines would of course have to be extended to cover the other potential counterexamples proposed in the literature (see Nolan 2017: 29, 1997: 569 fn. 21, Van der Laan 2004: 271). But since we are dealing with fairly isolated and involved examples, it doesn’t seem unreasonable for our fictionalists to hope that the best overall account of counterpossibles will still include the general principles (SIC) and (SLIC) and be able to account for the alleged counterexamples in a way compatible with them.

Alternatively, our fictionalists can even allow that certain utterances of counterfactuals are best accounted for by evaluating them with respect to a similarity ordering that contravenes (SIC) and (SLIC). But they can insist that, while it is in principle possible for a context to induce such a non-standard similarity
ordering, these orderings are to be disregarded when it comes to evaluating those counterfactuals which they propose, via (Operator), as analysantia of statements involving the fiction-operator. Given the extreme context-sensitivity of counterfactuals in natural language, counterfactual analyses of many philosophically interesting concepts need to be understood as accompanied by some such injunction to ignore extravagant similarity orderings in order to have any chance at coming out correct. For instance, consider Lewis’s injunction to disregard non-standard similarity orderings that render back-tracking counterfactuals true when evaluating the counterfactuals he proposes as analysantia of causal statements (1973: 565 fn. 10, 1979: 455-9).6

4 Conclusion

I have developed and defended the counterpossible strategy in response to the CFA. I have argued that fictionalists can establish a Safety Result sufficient for their purposes with assumptions that are consistent with their fiction being metaphysically impossible. I have then shown that fictionalists can appeal to familiar ideas pertaining to counterfactual reasoning, in particular to the role laws play in our evaluation of counterfactuals, in order to show that this strategy neither forces them to make an unprincipled distinction between counterlogical and countermetaphysical reasoning, nor deprives their fiction of the required content.7

Funding

This work was supported by the DFG, Emmy Noether Research Group Ontology After Quine (Universität Hamburg, WO 1896/1-1).

6Neither strategy would appear very promising if violations of (SIC) and (SLIC) weren’t exceptions but a widespread phenomenon. It has recently been argued that certain views about the metaphysics of omissions (Bernstein 2016: 2581) and causation (Nolan 2017: 28-9) lead to widespread violations of (SIC). This, then, appears to be a limitation of the present strategy: it had better not be combined with either of the views in question. Given that these views, their merits notwithstanding, can hardly be said to be without alternative, the attraction of the present solution seems to survive this limitation.

7Earlier versions of this paper were presented at the Issues on the (Im)possible V Conference in Bratislava, the Joint Sessions 2017 in Edinburgh, as well as a Research Colloquium and a Workshop on Fictionalism in Hamburg in 2018. I would like to thank the audiences on each of these occasions for helpful discussion and in particular Zsofia Zvolenszky for giving a response to a version of this paper at the first mentioned conference. Many thanks are also due to fellow members of the Emmy Noether Research Group Ontology After Quine as well as to two anonymous referees for this journal for helpful comments and criticisms.
References


