Reflection: Vacuism and the Strangeness of Impossibility

Rohan French

This is a preprint of a paper appearing in Yitzhak Y. Melamed and Samuel Newlands (Eds) *Modality* OUP, 2024.

Consider the following two sentences:

1. If Sally were to square the circle, we would be surprised.
2. If Sally were to square the circle, we would not be surprised.

These two sentences are both examples of counterfactual conditionals with impossible antecedents (henceforth 'counterpossible conditionals'). According to the orthodox semantics for counterfactuals due to Lewis (1973) and Stalnaker (1968) among others, what is required for a sentence like (1) to be true is for the most similar possible worlds to the actual world in which the antecedent is true (in this case, worlds in which Sally squared the circle), to be worlds in which the consequent is true (in this case, worlds in which we are surprised). According to these theories, because the antecedents of (1) and (2) are impossible, and thus true at no possible worlds, sentences (1) and (2) both come out vacuously true, as in all the possible worlds in which their antecedent is true (i.e. none of them), their consequent is also true. These are *vacuist* theories of counterpossible conditionals—theories according to which all counterpossibles are vacuously true.

Surely (1) is true and (2) is false, though! After all, if Sally had squared the circle she would have done something which is mathematically impossible, which is surely cause for surprise if anything is! According to *nonvacuist* theories of counterpossibles, conditionals like (1) and (2) have non-trivial truth conditions. Typically this is achieved by enriching the orthodox semantics by considering not only the most similar possible worlds where the antecedent is true, but instead considering the most similar possible *and impossible* worlds where the antecedent is true (as in Nolan (1997), for example). So in this case the nonvacuist will say that (1) is true because the most similar worlds, in this case impossible worlds, where Sally squares the circle are worlds where we are surprised, and (2) is false because in these most similar worlds in which Sally squares the circle it is not the case that we are not surprised.

By incorporating impossible as well as impossible worlds into their semantics nonvacuist theories are able express a range of fine grained distinctions between impossibilities, and it is this which allows them to count (1) as true and (2) as false. This results in nonvacuist logics invalidating a range of valid seeming inferences involving counterfactuals, depriving them of the most natural way of explaining the felt correctness of these inferences by having them come out logically valid like the vacuist can. As a result vacuists have charged nonvacuists of being committed to a counterfactual logic which is too weak, and that as a result they are unable to explain why those inferences are good inferences (cf. Williamson

(2007), p.174). For example, one of the most basic inference patterns involving counterfactuals involves the drawing out of deductive consequences in the scope of counterfactual suppositions. This pattern is captured by the rule CLOSURE below—informally, if A has B as a deductive consequence then for any sentence C, 'If it were that C, then it would be that A' has 'If it were that C, then it would be that B' as a deductive consequence. By classical logic "$2 + 2 = 4$ and it's not the case that $2 + 2 = 4$" has anything as a deductive consequence (this is the truth-functional inference principle *ex falso quodlibet*), and so 'we would not be surprised' as a consequence, and so by closure writing X for our sentence "$2 + 2 = 4$ and it is not the case that $2 + 2 = 4$" closure then gives us that 'If it were that X, then it would be that X' has 'If it were that X, then we would not be surprised' as a deductive consequence. But it's highly plausible that 'If it were that X, then it woudl be that X' is an instance of a theorem of counterfactual logic even by nonvacuist lights, while 'If it were that X, then we would not be surprised' seems to fall afoul of very similar issues to (2). So plausibly nonvacuists will need to give up principles like CLOSURE. Given that most of our use of counterfactuals does not involve counterpossibles on the face of it this seems to be a great cost to the nonvacuist.

How should the nonvacuist react to this situation? The common reaction is to put forward principles that 'tame' impossibility.[1] For example, the nonvacuist will point out that we only explicitly consider impossible situations when they are explicitly called by. So, for example, when I evaluate non-counterpossible conditionals like 'If Kangaroos were to not have tails, then they would fall over', I don't consider situations in which Kangaroos are inconsistent objects which both have and don't have tails. This line of thought motivates a constraint on similarity called in Nolan (1997) 'The Strangeness of Impossibility':[2]

> *Strangeness of Impossibility*: Possible worlds are more similar to the actual world than any impossible worlds is to the actual world.

Adapting the semantics given above, this means that to determine whether a conditional like 'If Kangaroos were to not have tails, then they would fall over' we consider the most similar worlds to the actual world, both possible and impossible, where the antecedent is true (i.e. worlds where Kangaroos don't have tails) and see whether the consequent is true there (i.e. whether they are worlds where Kangaroos fall over). On the (quite reasonable) assumption that it is metaphysically possible that Kangaroos don't have tails, the

---

[1] The suggestion offered here is a further elaboration of the general point made in section 3.1 of Berto et al. (2018) and section 12.3 of Berto and Jago (2013), the present argument making more explicit the manner in which we are able to recover the inferences validated by vacuist theories. See also Mares (1997), Theorem 6.3 of which is especially relevant.

[2] Nolan notes that he is 'hesitant to endorse' (Nolan (1997), p.566) the Strangeness of Impossibility in light of potential counterexamples to the principle (see Nolan (1997), p.551, 569), Vander Lan (2004),p.271 also gives another notable potential counterexample. We think that reasoning broadly along the lines given in Berto and Jago (2013), p.274 can diffuse all of these putative counterexamples, though, and so we dwell on them no further, instead expanding on the positive case for the Strangeness of Impossibility.

Strangeness of Impossibility tells us that he most similar worlds to the actual world where this is true must be possible worlds. This seems to point the way towards an account of how it is that the inferences which nonvacuists regard as strictly speaking invalid are nonetheless rationally compelling. What I will show here is a very clear and precise way in which this happens at the level of logics, allowing us to largely bypass contentious issues concerning the correct formal semantics for counterfactuals. In particular we will show how a very simple and natural nonvacuist logic can, in a certain sense, mimic a natural minimal vacuist logic modelled off the favored minimal logic for counterfactuals given in the Appendix of Williamson (2007).

***

Our formal language $\mathcal{L}$ will consist of denumerably many sentential variables $p_1, p_2, p_3, \ldots$, along with the unary connective '¬' (negation), '◇' (possibility), '□' (necessity), and the binary connectives '→' (the material conditional), and '>' (the counterfactual conditional). The connectives ∧, ∨ and ↔ are taken as defined out of these in the usual way. Throughout we will use uppercase Roman letters as schematic variables over arbitrary formulas of the language. We will consider as our prototypical vacuist counterfactual logic a slight variant on Williamson's minimal counterfactual logic, defined as follows, letting $\vdash_{\mathrm{Vac}}$ indicate theoremhood in our vacuist logic.[3]

K	If A is theorem of the modal logic K, then $\vdash_{\mathrm{Vac}} A$
MODUS PONENS	If $\vdash_{\mathrm{Vac}} A$ and $\vdash_{\mathrm{Vac}} A \to B$, then $\vdash_{\mathrm{Vac}} B$
NECESSITATION	If $\vdash_{\mathrm{Vac}} A$, then $\vdash_{\mathrm{Vac}} \Box A$
IDENTITY	$\vdash_{\mathrm{Vac}} A > A$
VACUITY	$\vdash_{\mathrm{Vac}} \Box B \to (A > B)$
CLOSURE	If $\vdash_{\mathrm{Vac}} (B_1 \wedge \ldots \wedge B_n) \to A$, then $\vdash_{\mathrm{Vac}} (C > B_1 \wedge \ldots \wedge C > B_n) \to (C > A)$
EQUIVALENCE	If $\vdash_{\mathrm{Vac}} C \leftrightarrow C^*$, then $\vdash_{\mathrm{Vac}} (C > A) \leftrightarrow (C^* > A)$

To see that this is a vacuist logic for counterfactuals, we will show that $\vdash_{\mathrm{Vac}} \neg\Diamond A \to (A > B)$—that it is a logical truth of Vac that all impossible propositions entail all counterfactuals with that proposition as their antecedent, regardless of their consequent.

(1)	$\neg\Diamond A \to \Box(A \to B)$	K
(2)	$\Box(A \to B) \to \big(A > (A \to B)\big)$	VACUITY

---

[3] In what follows we presume some familiarity with classical truth-functional logic and propositional modal logic. For more on modal logic the interested reader should consult the excellent Chellas (1980) and Hughes and Cresswell (1996). Note that Appendix 1 of Williamson (2007) is concerned with investigating the definition of the modal operators in terms of the counterfactual conditional, and so in place of our VACUITY he has $(\neg A > A) \to (A > B)$, where $(\neg A > A)$ is one of his candidate definitions of '□' in terms of '>'.

(3) $\neg\Diamond A \to \big(A > (A \to B)\big)$ 1,2, K, MODUS PONENS

(4) $\big((A \to B) \land A\big) \to B$ K

(5) $\Big(\big(A > (A \to B)\big) \land (A > A)\Big) \to (A > B)$ 4, CLOSURE

(6) $\big(A > (A \to B)\big) \to (A > B)$ 5, IDENTITY, K, MODUS PONENS

(7) $\neg\Diamond A \to (A > B)$ 3,6, K, MODUS PONENS

***

As we saw above, if we assume the Strangeness of Impossibility principles, such as CLOSURE and EQUIVALENCE, have a number of instances which are safe for the nonvacuist, namely those instances where the antecedents of counterfactuals involved are possible—the Strangeness of Impossibility having as a consequence that in such cases we only need to consider possible worlds and so in such special cases things will proceed just as the vacuist claims they should in all cases. This motivates the folllowing rules, where $\vdash_{\text{NVac}}$ indicates theoremhood in the Strangeness of Impossibility endorsing nonvacuists logic:

- $\Diamond$-CLOSURE: If $\vdash_{\text{NVac}} (B_1 \land \ldots \land B_n) \to A$, then $\vdash_{\text{NVac}} \Diamond C \to \big((C > B_1 \land \ldots \land C > B_n) \to (C > A)\big)$
- $\Diamond$-EQUIVALENCE: If $\vdash_{\text{NVac}} C \leftrightarrow C^*$, then $\vdash_{\text{NVac}} \Diamond C \to \big((C > A) \leftrightarrow (C^* > A)\big)$

Furthermore, the sentence characteristic of the Strangeness of Impossibility itself corresponds to the safe version of VACUITY

- SIC: $\vdash_{\text{NVac}} \Diamond A \to \big(\Box B \to (A > B)\big)$

Note that it is this principle which semantically corresponds (in the sense of van Benthem (1984)) to the Strangeness of Impossibility (as was pointed out in French et al (2022), p.265) and not the principle $(A > B) \to (\Diamond A \to \Diamond B)$ as was claimed in Williamson (2017), p.200, and strongly implied in Berto et al. (2018), p.667.[4]

Combining these with the parts of Vac which are unobjectionable to the nonvacuist we get the logic $\vdash_{\text{NVac}}$, which is defined as follows:

K	If A is theorem of the modal logic K, then $\vdash_{\text{NVac}} A$
MODUS PONENS	If $\vdash_{\text{NVac}} A$ and $\vdash_{\text{NVac}} A \to B$, then $\vdash_{\text{NVac}} B$
NECESSITATION	If $\vdash_{\text{NVac}} A$, then $\vdash_{\text{NVac}} \Box A$
IDENTITY	$\vdash_{\text{NVac}} A > A$

[4] In the setting of Nolan (1997), which Williamson explicitly mentions, this principle corresponds not to the Strangeness of Impossibility, but to what Nolan (1997), p.566 calls *The Lesser Strangeness of Impossibility*, which informally states that 'no impossible world is less distant than any possible world'.

SIC	$\vdash_{\mathrm{NVac}} \Diamond A \to (\Box B \to (A > B))$	
$\Diamond$-CLOSURE	If $\vdash_{\mathrm{NVac}} (B_1 \wedge \ldots \wedge B_n) \to A$, then $\vdash_{\mathrm{NVac}} \Diamond C \to ((C > B_1 \wedge \ldots \wedge C > B_n) \to (C > A))$	
$\Diamond$-EQUIVALENCE	If $\vdash_{\mathrm{NVac}} C \leftrightarrow C^*$, then $\vdash_{\mathrm{NVac}} \Diamond C \to ((C > A) \leftrightarrow (C^* > A))$	

Importantly NVac does not validate $\neg\Diamond A \to (A > B)$. To see this we can use a toy model from French et al (2022), p.268 which consists of only two worlds, a possible world x and a single impossible world y where, approximately speaking, the most similar world to our possible world is that possible world itself, and so for a counterfactual $A > B$ with a possible antecedent (and so A true at our sole possible world x) to be true at x we must have B true at x also. Our logic NVac is essentially the logic dubbed Successful-Quasi-Vacuism in French et al (2022), restricted to the present language,[5] and as a result every theorem of $\vdash_{\mathrm{NVac}}$ is true at x in the model for Successful-Quasi-Vacuism given there. As noted there $\neg\Diamond A \to (A > B)$ can be made false in such a model, for suppose A is true at our impossible world only, and B is true nowhere. Then we have $\neg\Diamond A$ true at x, but $A > B$ false, as the nearest A-world (our impossible world y) fails to be a B-world. So as $\neg\Diamond A \to (A > B)$ is not true in that model at x, while every theorem of $\vdash_{\mathrm{NVac}}$ is, it follows That $\neg\Diamond A \to (A > B)$ is not a theorem of $\vdash_{\mathrm{NVac}}$—the logic $\vdash_{\mathrm{NVac}}$ is properly nonvacuist.

One thing which is important to note at this point, which we will need to appeal to in a moment, is the fact that $\vdash_{\mathrm{NVac}} \subseteq \vdash_{\mathrm{Vac}}$, as all the axioms of NVac are either axioms of Vac or the truth-functional weakenings of axioms of Vac, and anything which follows from theorems of NVac by $\Diamond$-CLOSURE or $\Diamond$-EQUIVALENCE also follows from CLOSURE or EQUIVALENCE along with truth-functional reasoning in K.

***

Recall now the motivating idea behind the Strangeness of Impossibility, namely that we only consider impossibilities when evaluating counterfactuals when impossibilities are explicitly involved. If they accept the Strangeness of Impossibility, and consequently a logic like NVac, the nonvacuist is in a position to understand vacuist talk as being correct, but about a subtly different subject matter—the nonvacuist is able to agree with the vacuist if they reinterpret their talk of counterfactual conditionals as talk of counterfactual conditionals conditional on the possibility of their antecedent. So, for example, the nonvacuist can understand the vacuists (by their lights, incorrect) claim that (2) is true, as the (by their lights, correct) claim that (3) is true:[6]

---

[5] The rules $\Diamond$-CLOSURE and $\Diamond$-EQUIVALENCE are not covered there, but are easily seen to preserve truth at x in that model.

[6] The argument pursued here is reminiscient of one used by Salmon (1989) in the course of arguing against the correctness of the modal logic S5 as the correct logic for metaphysical necessity. Salmon argues that the strongest acceptable logic for metaphysical necessity is the weak modal logic KT, and that philosophers advocating for S5 conflate ‘possibility’ and

3. Either it is impossible for Sally to square the circle, or if Sally were to square the circle, we would not be surprised.

Put more formally we are able to translate between vacuist theories and appropriate SIC-endorsing nonvacuist theories using the following translation τ which maps formulas of $\mathcal{L}$ to formulas of $\mathcal{L}$.

- $t(p) = p$
- $t(\neg A) = \neg t(A)$
- $t(A \to B) = t(A) \to t(B)$
- $t(\Diamond A) = \Diamond t(A)$
- $t(\Box A) = \Box t(A)$
- $t(A > B) = \Diamond t(A) \to \big(t(A) > t(B)\big)$

It is easy to show by induction on length of derivations that if $\vdash_{\mathrm{Vac}} A$ then $\vdash_{\mathrm{NVac}} t(A)$. For the basis case the only interesting cases are when A is an instance of IDENTITY or VACUITY, in which case we have $\vdash_{\mathrm{NVac}} t(A)$ by IDENTITY and truth-functional reasoning in K, or SIC and truth-functional reasoning in K respectively. For the induction step, suppose that for formulas B with derivations of length $< n$ in $\vdash_{\mathrm{Vac}}$ we have $\vdash_{\mathrm{NVac}} B$, and suppose (for example) that the next inference applied is CLOSURE. Then we have $\vdash_{\mathrm{Vac}} (B_1 \wedge \ldots \wedge B_n) \to A$ and so by our induction hypothesis $\vdash_{\mathrm{NVac}} t\big((B_1 \wedge \ldots \wedge B_n) \to A\big)$, i.e. $\vdash_{\mathrm{NVac}} \big(t(B_1) \wedge \ldots \wedge t(B_n)\big) \to t(A)$. Then by $\Diamond$-CLOSURE we have

$$\vdash_{\mathrm{NVac}} \Diamond t(C) \to \Big(\big(t(C) > t(B_1) \wedge \ldots \wedge t(C) > t(B_n)\big) \to \big(t(C) > t(A)\big)\Big)$$

But truth-functional logic, and hence K, contains all instances of the formula $A \to (B \to C) \to \big((A \to B) \to (A \to C)\big)$, and so K and Modus Ponens applied to the above yields:

$$\vdash_{\mathrm{NVac}} \Diamond t(C) \to \big(t(C) > t(B_1) \wedge \ldots \wedge t(C) > t(B_n)\big) \to \Diamond t(C) \to \big(t(C) > t(A)\big).$$

Making use of the truth-functional tautology $\big(A \to (B \wedge C)\big) \leftrightarrow \big((A \to B) \wedge (A \to C)\big)$ and Modus Ponens applied to the above then gives:

$$\vdash_{\mathrm{NVac}} \Big(\Diamond t(C) \to \big(t(C) > t(B_1)\big) \wedge \ldots \wedge \Diamond t(C) \to \big(t(C) > t(B_n)\big)\Big) \to (\Diamond t(C) \to (t(C) > t(A))),$$

which is to say, $\vdash_{\mathrm{NVac}} t\big((C > B_1 \wedge \ldots \wedge C > B_n) \to (C > A)\big)$.

'necessity' with 'actual possibility' and 'actual necessity'. This is made precise by Williamson (1998), p.98f in the form of the following translation result: A is a theorem of S5 then its translation $\tau(A)$ is a theorem of the logic KT@S (KT enriched with a rigidifying actuality operator '@') where $\tau(\Diamond A) = @\Diamond\tau(A)$, and similarly for $\Box$.

Moreover, the theorems of Vac are *precisely* the theorems of NVac of the form $t(\cdot)$. To see this first note that we can prove by induction on the complexity of A that $\vdash_{\text{Vac}} A \leftrightarrow t(A)$. For the case where A is of the form $B > C$ we make use of the fact that $\vdash_{\text{Vac}} \big(\Diamond A \rightarrow (A > B)\big) \rightarrow (A > B)$, this following from the provability of $\neg\Diamond A \rightarrow (A > B)$ (shown above) and the truth-functional tautology $(\neg A \rightarrow C) \rightarrow ((A \rightarrow C) \rightarrow C)$, the other direction of the equivalence being a truth-functional tautology. Now, suppose that $\vdash_{\text{NVac}} t(A)$. Then as $\vdash_{\text{NVac}} \subseteq \vdash_{\text{Vac}}$ we also have $\vdash_{\text{Vac}} t(A)$ and so by truth-functional reasoning and the fact that $\vdash_{\text{Vac}} A \leftrightarrow t(A)$ we have $\vdash_{\text{Vac}} A$ as desired.

***

Accepting the Strangeness of Impossibility allows the nonvacuist to reply to the complaint that they cannot explain the correctness of inferences involving counterfactuals. Whenever the vacuist claims that a given inference is valid the nonvacuist translation of that inference will also be valid. In many contexts that will amount to that inference being truth-preserving if it's clear in that context that the antecedents of the conditionals are possible. This is, I think, the clearest positive argument in favor of the Strangeness of Impossibility.

References

van Benthem, J. (1984). Correspondence Theory, In D, Gabbay. and F. Guenthner (Eds.), *Handbook of Philosophical Logic*. Synthese Library, vol 165. pp. 167-247. Dordrecht: Springer

Berto, F., R. French, G. Priest, and D. Ripley (2018). Williamson on Counterpossibles. *The Journal of Philosophical Logic* 47, 693-713.

Berto, F. and M. Jago (2013). *Impossible Worlds*. Oxford: Oxford University Press.

Chellas, B. (1980). *Modal Logic: An Introduction*. Cambridge: Cambridge University Press.

French, R., P. Girard, and D. Ripley (2022). Classical Counterpossibles. *The Review of Symbolic Logic* 15(1), 259-257.

Hughes, G. and M. J. Cresswell (1996). *A New Introduction to Modal Logic*. London: Routledge.

Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.

Mares, E. D. (1997). Who's afraid of impossible worlds? *Notre Dame Journal of Formal Logic* 38, 516-526.

Nolan, D. (1997). Impossible worlds: a modest approach. *Notre Dame Journal of Formal Logic* 38, 535-572.

Salmon, N. (1989). The Logic of What Might Have Been. *The Philosophical Review* 98(1), 3-34.

Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in Logical Theory*, pp. 98-112. Oxford: Blackwell.

Vander Laan, D. (2004). Counterpossibles and similarity. In F. Jackson and G. Priest (Eds.), *Lewisian Themes: The Philosophy of David K. Lewis*, pp. 258-275. Oxford: Oxford University Press.

Williamson, T. (1998). Iterated attitudes. *Proceedings of the British Academy* 95, 85-133.

Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Oxford University Press.

Williamson, T. (2017). Counterpossibles in Semantics and Metaphysics. *Argumenta* 2(2), 195-226.