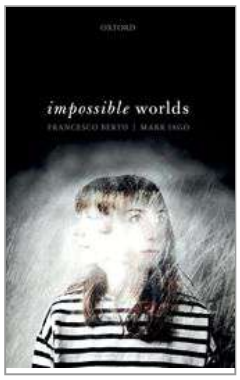


Impossible Worlds
Francesco Berto and Mark Jago

Print publication date: 2019
Print ISBN-13: 9780198812791
Published to Oxford Scholarship Online: August 2019
DOI: 10.1093/oso/9780198812791.001.0001

Counterpossible Conditionals

Francesco Berto
Mark Jago
Rohan French
Graham Priest
David Ripley

DOI:10.1093/oso/9780198812791.003.0012

Abstract and Keywords

Vacuism is the view that all counterpossibles are trivially true. There are reasons to think it incorrect. An impossible worlds semantics for counterfactuals is offered, which makes room for non-trivial counterpossibles. One principle which pins down its application is the *Strangeness of Impossibility* condition: for any given possible world, any impossible worlds is further away from it than any possible world is. A number of Williamson's objections to the non-vacuist approach are discussed and it is argued that they can be overcome. The question of whether counterfactuals in general should permit the substitution of rigidly coreferential terms is then raised. Having defended non-vacusim against Williamson's objections, a range of arguments in its favour are considered.

Keywords: counterpossibles, vacuism, Strangeness of Impossibility condition, substitution, Williamson

12.1 Why Counterpossibles?

One of the most discussed applications of impossible worlds has to do with the treatment of counterpossible conditionals. These are counterfactuals whose antecedent is true at no possible world. As you may recall from §1.3, the Lewis-Stalnaker semantics has it that, if there are no A-worlds, $A \boxrightarrow B$ comes out automatically true. The conditional with the same antecedent and opposite consequent, $A \boxrightarrow \neg B$, comes out true, too, for the same reason. In general, all

counterpossibles are vacuously true. The standard treatment of counterfactuals implies *vacuism* about counterpossibles.

To many, including Brogaard and Salerno (2013), Bernstein (2016), Bjerring (2014), Krakauer (2012), Nolan (1997), and Priest (2008), vacuism seems wrong (§1.3). These authors have come up with numerous examples of counterfactuals with impossible antecedents, such that the consequent matters for the truth value of the whole. Recall Nolan's (1997) pair of Hobbes-sentences from §1.3:

(p.268)

(1.18) If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would have cared.

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

The intuition is that Hobbes's squaring the circle would have made no difference with respect to the life of those sick children. The second Hobbes-sentence should, then, be true for this reason, and for the same reason, the first Hobbes-sentence should be false. (Bernstein (2016) gives a similar argument.)

Other examples of non-vacuous counterpossibles arise with non-causal notions of 'making a difference'. Anna and her singleton, {Anna}, are modally inseparable: necessarily, one exists just in case the other does. Yet we can make good sense of the idea that a particular set's existence depends on a general framework of sets, in a way that Anna's existence doesn't.

(12.1) If there hadn't been any sets, {Anna} wouldn't have existed.

is true, whereas

(12.2) If there hadn't been any sets, Anna wouldn't have existed.

is false. For whether or not sets exist makes no difference to Anna's existence.

Brogaard and Salerno (2013) propose that counterpossibles such as these can help in the analysis of a thing's *essence*. They agree with Fine's (1994) idea that {Anna} is not involved in Anna's essence, even though the two are modally inseparable. They argue that we can explain this using the difference in truth-value between (12.1) and (12.2). We agree that there's a link between essence and counterpossibles such as these, but we're not so sure about Brogaard and Salerno's direction of explanation. Couldn't it be that (12.2) is false because Anna's essence doesn't involve any sets? If so, it may be that essences play a role in explaining counterfactuals (including counterpossibles), and not vice versa.

(p.269) The rest of this chapter will be largely structured as a dialogue with Timothy Williamson, who presents a series of powerful objections to non-vacuism (Williamson 2007, 2010, 2017). Discussing these objections will give us the opportunity to delve into the details of a non-vacuist theory of counterfactuals with impossible worlds. In §§12.3–12.5, we'll discuss three arguments against non-vacuist semantics and, in §12.6, we'll discuss Williamson's attempts to undermine the intuitive pull of non-vacuism.

12.2 A Semantics for Counterpossibles

The obvious way to free the Lewis-Stalnaker semantics from vacuism is to expand it by adding impossible worlds. Start with a standard propositional language $\mathscr{L}$ like the one of §4.1 and add our counterfactual conditional $\Box\!\!\rightarrow$, so that if A and B are formulas, then so is $A \Box\!\!\rightarrow B$.

A *frame* $\mathscr{F}$ is now a triple $\langle\!\langle\, W, N, \{R_A \mid A \in \mathscr{L}\} \,\rangle\!\rangle$, with W the set of worlds, $N \subseteq W$ the subset of normal (possible) worlds, and each R_A an accessibility relation on W (one for each formula in the language). We read '$R_A w w_1$' as meaning that w_1 is *ceteris paribus* like w, but A is true at w_1. (For this reading to make sense, we'll need an extra constraint on each R_A; see below.)

A frame becomes a model $\mathscr{M} = \langle\!\langle\, W, N, \{R_A \mid A \in \mathscr{L}\}, v \,\rangle\!\rangle$, when endowed with a valuation function v assigning truth values (0 or 1) to atoms at worlds in N and to all formulas at worlds in $W - N$. (So as before, impossible worlds are worlds where complex formulas are treated as atomic.) The truth conditions for the operators other than $\Box\!\!\rightarrow$ at $w \in N$ are as in §4.1. For simplicity, we do without the accessibility relation for $\Box$ and $\Diamond$, which we treat as unrestricted universal and existential quantifiers over possible worlds. As for the counterfactual:

(S$\Box\!\!\rightarrow$) $v_w(A \Box\!\!\rightarrow B) = 1$ if for all w_1 such that $R_A w w_1$, $v_{w_1} B = 1$, and 0 otherwise.

(p.270) Logical truth and validity are, respectively, truth and truth preservation at all normal worlds in all models. This gives us classical **S5** modal logic for the extensional connectives and $\Box$ and $\Diamond$. The only operator that looks at impossible worlds is $\Box\!\!\rightarrow$. With no constraints on the accessibility relations R_A, we have a basic system of conditional logic.

Stronger systems can be obtained, as usual, by adding constraints on the accessibility relations. Their intended understanding clearly motivates the following:

(12.3) If $R_A w w_1$ then $v_{w_1}(A) = 1$

(12.4) If $v_w(A) = 1$ then $R_A w w$

The former says that A is true at all R_A-accessible worlds. The latter says that, if A is true at w, then nothing is closer to w than itself. This corresponds to what Lewis (1973b) called 'Weak Centring'.

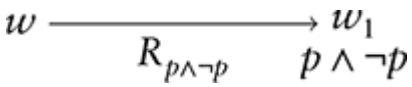
These conditions have an effect only when w is a possible world, since the R_As are not involved in determining the truth value of anything at an impossible world. They guarantee, respectively, that '$\Box\!\!\rightarrow$' satisfies counterfactual self-implication and *modus ponens*:

(12.5) $\vDash A \Box\!\!\rightarrow A$

(12.6) $A, A \Box\!\!\rightarrow B \vDash B$

These inferences are clearly desirable for the counterfactual conditional.

The semantics is non-vacuist. To see this, consider this model, with $N = \{w\}$:



At w, $p \wedge \neg p \Box\!\!\rightarrow q$ is false, even though the antecedent is contradictory. For w can access the impossible world w_1 (via $R_{p \wedge \neg p}$), where q is not true, even though $p \wedge \neg p$ is.

(p.271) 12.3 The Strangeness of Impossibility Condition

In order for the R_As genuinely to express world similarity, we would need to impose a comparative similarity relation on worlds, or a 'system of spheres', expanding Lewis (1973b)'s approach with the addition of impossible worlds. We will not set things up in this way. The problem of how similarity should work when impossible worlds are around is a tricky one. Some of the issues are orthogonal to the topics we are to discuss in this chapter. However, one further constraint on the R_As will play an important role in our discussion, the *Strangeness of Impossibility Condition*:

(SIC) If $v_w A = 1$ for some $w \in N$ and $R_A w w_1$, then $w_1 \in N$.

If A is true at some possible world w, which looks via R_A at w_1, then w_1 is possible, too. The thought expressed by the constraint is the *prima facie* plausible one that, to evaluate the truth at a possible world of a conditional with a possible antecedent, we never look at impossible worlds. Thinking in terms of closeness between worlds, the condition says that any possible world is closer to a possible world w than any impossible world is. Impossible worlds are kept at a distance for as long as they can be: they're *strange*. (Hence the name, due to Nolan (1997). Jago (2014a) and Mares (1997) also endorse the approach.)

With (SIC) in place, it is easily checked that our semantics validates:

(12.7) $\Diamond A, A \Box\!\!\rightarrow B \vDash \Diamond B$

It has further important consequences for validity, connected to an objection raised by Williamson (2007) over an impossible worlds logic for counterfactuals:

> We may wonder what logic of counterfactuals [non-vacuists] envisage. If they reject elementary principles of the pure logic of counterfactual conditionals, that is an unattractive feature of their position.
>
> (Williamson 2007, 174)

(p.272) Williamson does not make explicit which logic he has in mind as 'the pure logic of counterfactual conditionals', or which of its principles are 'elementary'. But he makes use of a weak counterfactual logic, presented proof-theoretically (2010, 85). We assume that the distinctively counterfactual axioms and rules of this system give a sense of what Williamson means. We will consider three (using '$\vdash$' for theoremhood and '$\leftrightarrow$' for material equivalence):

(12.8) $\vdash A \mathrel{\Box\!\!\rightarrow} A$

(12.9) If $\vdash A \leftrightarrow B$ then $\vdash (A \mathrel{\Box\!\!\rightarrow} C) \leftrightarrow (B \mathrel{\Box\!\!\rightarrow} C)$

(12.10) If $\vdash B_1 \wedge \cdots \wedge B_n \supset C$ then $\vdash (A \mathrel{\Box\!\!\rightarrow} B_1) \wedge \cdots \wedge (A \mathrel{\Box\!\!\rightarrow} B_n) \supset (A \mathrel{\Box\!\!\rightarrow} C)$

Of these, our semantics verifies only (12.8). So should (12.9) and (12.10) be endorsed?

The former has it that whenever A and B are provably equivalent, then so too are $A \mathrel{\Box\!\!\rightarrow} C$ and $B \mathrel{\Box\!\!\rightarrow} C$. If the extensional fragment of the logic is classical (as in our semantics), then any classical contradiction is provably equivalent to any other. So (12.9) implies that $p \wedge \neg p \mathrel{\Box\!\!\rightarrow} C$ is provably equivalent to $q \wedge \neg q \mathrel{\Box\!\!\rightarrow} C$, for any choice of p, q, and C. In particular, $p \wedge \neg p \mathrel{\Box\!\!\rightarrow} q \wedge \neg q$ is provably equivalent to $q \wedge \neg q \mathrel{\Box\!\!\rightarrow} q \wedge \neg q$. But the latter is provable, given (12.9), and hence so is the former:

$$\vdash p \wedge \neg p \mathrel{\Box\!\!\rightarrow} q \wedge \neg q$$

Quite generally, (12.8) and (12.9) imply that any provably contradiction counterfactually implies any other. But why think this? If Graham Priest really had found a box that's both empty and not empty (as in his story, §11.3), would it really be both raining and not raining in Amsterdam? We don't think that's plausible. Since we reject the conclusion, but find (12.8) hard to deny, we reject (12.9).

Similar problems arise in connection with (12.10). Classically we have $\vdash (p \wedge \neg p) \supset q$. From (12.10), we infer

$$\vdash ((p \wedge \neg p) \mathrel{\Box\!\!\rightarrow} (p \wedge \neg p)) \supset ((p \wedge \neg p) \mathrel{\Box\!\!\rightarrow} q)$$

(p.273) from which, using (12.8), we obtain:

$$\vdash (p \wedge \neg p) \boxdot\!\!\rightarrow (p \wedge \neg p)$$

This gives us $\vdash (p \wedge \neg p) \boxdot\!\!\rightarrow q$. If we accept all that, then any contradiction will counterfactually imply anything at all. But why think this? It's wrong to think that, were it raining and not raining, giraffes would stand on their horns. So one of (12.8), (12.10), and classical logic must go. We reject (12.10).

Williamson's principles combine to yield bad predictions about counterfactuals with contradictory antecedents. Counterfactual suppositions can take us beyond logical bounds; they can lead us to entertain situations in which logically equivalent claims come apart, or in which a claim can hold without all its consequences holding. For non-vacuists, these are not 'unattractive features' of their view; rather, they provide one of the main intuitive motivations for it. Of course, such intuitions can be challenged: we will come to this in §12.6. But simply assuming that they are wrong would be dialectically unhappy.

Non-vacuists should reject (12.9) and (12.10). There are closely related principles they may accept, however:

(12.11) If $\vdash A \leftrightarrow B$ then $\Diamond A \vdash (A \boxdot\!\!\rightarrow C) \leftrightarrow (B \boxdot\!\!\rightarrow C)$

(12.12) If $\vdash B_1 \wedge \cdots \wedge B_n \supset C$ then $\Diamond A \vdash (A \boxdot\!\!\rightarrow B_1) \wedge \cdots \wedge (A \boxdot\!\!\rightarrow B_n) \supset (A \boxdot\!\!\rightarrow C)$

These are just like (12.9) and (12.10), except that the validities they yield have as a premise that a certain claim is possible.

Our semantics validates $\Diamond(A \wedge \neg A) \vdash B$ for any A and B, and so the arguments above against (12.9) and (12.10) do not extend to (12.11) and (12.12). With (SIC) in place, we never have to go outside the domain of possible worlds to evaluate an inference, so long as the antecedents of all the conditionals we are dealing with are possible. As a result, all the valid inferences of merely-possible-world semantics (including (12.9) and (12.10)) are recoverable enthymematically, simply by adding suppressed premises of the form $\Diamond A$ (as in (12.11) and (12.12)). In that sense, adding impossible worlds loses us nothing.

(p.274) We get a lot from accepting (SIC). But is it acceptable? Some authors (Bernstein 2016, Nolan 1997, Vander Laan 2004) have argued against it. In particular, counterexamples have been proposed to (12.7), which follows from (SIC). Nolan (1997, 2017) offers these:

> (12.13) If intuitionistic logic came to be thought of as a much more satisfactory basis for mathematics by experts, and if intuitionistic investigations led to breakthroughs in many areas, … then intuitionistic logic would turn out to be correct after all. (Nolan 1997, 550)

(12.14) If Gödel had believed Fermat’s Last Theorem to be false, it would have been. (Nolan 1997, 569)

(12.15) If the bag had 63 balls in it, 63 would have been a square number. (Nolan 2017, 17)

Each of these conditionals has the right form to be a counterexample to (12.7): a possible antecedent and an impossible consequent. But are they true? The context for (12.14) involves a person in awe of Gödel’s ability, who thinks that whatever was believed by Gödel in mathematics must be true. It seems to us that intuitionistic logic would not turn out to be correct even if most experts agreed on its value, and that Fermat’s Last Theorem would stay true, even if Gödel had believed otherwise.

For (12.15), the context is one in which a person teaches a boy how square numbers work by arranging balls in a square grid, then putting them in a bag and counting the balls that come out. Sometimes the total is 16, sometimes 25, and so on. The conditional is then uttered by the person on an occasion where 63 balls are counted. In this case, we agree with Nolan (2017, 17) that (12.15) would be an appropriate thing to say. But it still does not sound literally true to us. Uttering (12.15) in that context seems to us way to convey the thought that some miscounting must have taken place. In that respect, it’s like ‘if Trump were smart, I’d be a monkey’s uncle’, whose antecedent is possibly true while its consequent looks **(p.275)** like a metaphysical impossibility. We don’t utter such things as a commitment to their literal truth.

One way to motivate (SIC) is by analogy with what Bennett (2003, 227) calls *counterlegal* conditionals. (We thank Jorge Ferreira for calling our attention to this point.) When we evaluate ordinary counterfactuals, we look at worlds like the world of evaluation, up to or around the time of the antecedent, and which are nomologically possible (Bennett 2003, 198). So we already have a ‘Strangeness of Nomological Impossibility Condition’ in play in the evaluation of ordinary counterfactuals. Nomologically possible worlds form a sphere, in that they are closer to the base world than nomologically impossible ones. So when we evaluate ordinary counterfactuals whose antecedents comply with our laws, we never look beyond nomologically possible antecedent-worlds.

Things are different when we deal with *counterlegals*, whose antecedents are causally or nomologically impossible: ‘if gravity obeyed an inverse cube law, then our months would be shorter’. Then we need to move beyond the nomologically possible, and look at the antecedent-worlds that are nomologically most similar to the base world, despite breaking some of its causal laws. (While Bennett doubts that there is any principled way to do it, we are more optimistic. Thought experiments in the natural sciences often have us suppose situations which violate actual physical laws, often with widespread consensus. This suggests that

there is a principled way to evaluate the corresponding counterfactual formulations.)

Analogously, we claim that something like (SIC) is in play with counterfactuals whose antecedents do not violate a law which is absolutely necessary. Possible worlds form a sphere, in that they are closer to the base world than impossible ones. We only look at impossible worlds when the antecedent forces us to move outside the sphere of absolute possibility. We do it when engaging in philosophical or logical thought experiments, as when we counterfactually suppose a logical or mathematical theory we deem (necessarily) wrong, in order to draw unpalatable consequences from it, by way of *reductio*. (We will come back to *reductio* reasoning in §12.5.)

(p.276) 12.4 Substitutivity of Identicals

What does our approach to counterpossibles say about identity and the substitution of rigidly coreferential terms? To investigate the issue, we extend our language with n-ary predicates for each n, the two-place identity predicate, '=', and a set of individual constants, with the usual rules of well-formedness. In particular, if 'a' and 'b' are any constants, then '$a = b$' is an atomic formula. (Extending the semantics to the quantifiers is a non-trivial matter, due to the presence of impossible worlds. We need not go into details here, for they are not germane to what follows; but see Priest's 'matrix semantics' (2008, chapters 18 and 23).)

Models now contain a domain and an interpretation function, assigning an element of the domain to each constant, a subset of the domain to each monadic predicate, and (for $n > 1$) an n-tuple to each n-ary predicate. For atomic sentences other than atomic sentence letters and worlds $w \in N$, v_w is defined in the usual way, in terms of the interpretation. In particular, an atomic identity statement '$a = b$' is true iff 'a' and 'b' denote the same element of the domain. (As before, when $w \in W - N$, v_w treats all sentences as atomic.)

As a consequence, v always satisfies these constraints when $w \in N$:

(12.16) $v_w(a = a) = 1$

(12.17) For atomic A, if $v_w(a = b) = 1$ then $v_w(A) = v_w(A[b/a])$

(12.18) For any $w_1 \in N$, $v_{w_1}(a = b) = v_w(a = b)$

It is then easy to establish that, if A is any sentence in which a does not occur within the scope of a '$\Box\rightarrow$', $w \in N$, and $v_w(a = b) = 1$, then $v_w(A) = 1$ iff $v_w(A[b/a]) = 1$. So the *Substitutivity of Identicals*, as we will call it, holds in such contexts.

As there are no constraints on v at impossible worlds, *Substitutivity of Identicals* does not hold for impossible worlds, just as we would expect. As a consequence,

Substitutivity of Identicals is not valid when substitution is within the scope of counterfactuals, for counterfactuals **(p.277)** may look to impossible worlds. Again, that's to be expected when impossible antecedents are around. For example,

> (12.19) If Hesperus were not Phosphorus, then modern astronomy in particular would be badly mistaken.

That's true, and Hesperus is Phosphorus; yet it's not the case that

> (12.20) If Phosphorus were not Phosphorus, then modern astronomy in particular would be badly mistaken.

Rather, it would be (mainstream) modern logic in particular that is badly mistaken.

Now consider the following pair, from Williamson (2007, 174–6):

> (12.21) If Hesperus had not been Phosphorus, Phosphorus would not have been Phosphorus.

> (12.22) If Hesperus had not been Phosphorus, Hesperus would not have been Phosphorus.

We take the appropriate evaluation of these to be as follows: (12.21) is false and (12.22) is true. Had Hesperus not been Phosphorus, nothing would have followed about the self-identity of Hesperus or Phosphorus. This seems to suggest that surrendering metaphysical truths in a counterfactual supposition does not force us away from logical truths concerning the same subject matter. Our semantics agrees on this. As an instance of (12.5), (12.22) is valid, whereas the fact that Hesperus is Phosphorus does not imply (12.22). In general, on our semantics, $a = b$ does not entail $a \neq b \mathrel{\Box\!\!\rightarrow} a \neq a$.

Substitutivity of Identicals can fail on our semantics only when the substitution in question is within the scope of a counterfactual. Counterfactuals create hyperintensional contexts. Our counterfactuals are sensitive to distinctions between impossibilities, which are invisible in a standard intensional framework using possible worlds.

Yet Williamson (2007, 175) finds this 'highly implausible'. The reason for this has two premises. Hyperintensionality, he claims, occurs **(p.278)** only in constructions that are 'about representational features', such as epistemic and intentional contexts. But, he adds, counterfactuals are not about representational features in this way.

There is reason to doubt each of Williamson's premises. First, one might think, with Lycan (2001), that counterfactuals do involve epistemic features (and for reasons wholly independent of the non-vacuism debate). If so, then Williamson's argument falls apart. The failure of substitution in (12.21) and (12.22) would be of a piece with the failure that occurs when we note that it is a priori that Phosphorus is Phosphorus, but not that Hesperus is Phosphorus. (Brogaard and Salerno (2013, 654) appeal directly to the alleged epistemic aspect of counterfactuals to explain failures of substitutivity like this.) One might even agree with Thomasson (2007) that metaphysical modality itself involves representational elements, even when carefully contrasted with epistemic modalities.

That's all rather controversial, and one may not want to commit to any of those views. But even if we decide against them all, we should question the other Williamsonian premise. An operator's being hyperintensional does not entail its being representational or broadly epistemic. According to Nolan (2014), there are hyperintensional contexts that are not 'about representational features', and counterfactuals may well be among these. *Metaphysical grounding* is often taken to be a wholly worldly, non-representational, but hyperintensional concept (Correia and Schnieder 2012, Fine 2012b). If disjuncts ground disjunctive states of affairs or propositions (as many grounding theorists suppose), then *that A* is a ground for *that A* ∨ (*A* ∧ *B*), but not *vice versa*. So grounding is hyperintensional, since *A* ∨ (*A* ∧ *B*) is logically equivalent to *A*. Wilson (2018) argues that non-vacuism follows from a counterfactual approach to grounding.

Similarly, *essence* is a hyperintensional metaphysical concept which is frequently taken to be wholly non-representational. We can derive 'Anna exists' from '{Anna} exists' and vice versa, and yet what's essential to Anna'a existence differs from what's essential to {Anna}'s existence. In particular, {Anna}'s essence depends on Anna, whereas Anna's essence doesn't depend on {Anna} (Fine 1994). It seems **(p.279)** essential that the state of affairs *that it's raining or not raining* somehow involve the state of affairs *that it's raining*. But that state of affairs isn't essential to *that Trump will be impeached or he won't*. These logically equivalent complex states of affairs have different essences, and yet there's nothing representational or epistemic to them or their essences.

To see how counterfactuals might be hyperintensional without being about representations, simply return to the semantics we sketched above. Assume, together with Barcan, Kripke (1971), and Williamson (2007, 161), that if *a* = *b*, then it is necessary for *a* to be *b*. Notice that our semantics above conforms to this: the truth values of identity statements '*a* = *b*' do not change across possible worlds. Then *a*'s not being *b* is a way things just cannot be. Worlds at which *a* is not *b* are impossible worlds. There need be nothing epistemic about this, any more than there is about a world which hosts a physical impossibility, such

(supposing Einstein was right) as something travelling faster than the speed of light.

One may grant that Williamson's argument about 'representational features' is problematic, but still think that counterfactuals allow for substitution of identicals. Williamson (2007, 174) bolsters this impression with the following argument:

> (12.23) If the rocket had continued on its course, it would have hit Hesperus.
>
> (12.24) Hesperus = Phosphorus
>
> (12.25) Therefore, if the rocket had continued on its course, it would have hit Phosphorus.

This, Williamson claims, is 'unproblematically valid' (2007, 174).

We agree that the argument steps are truth-preserving, but deny that this is so in virtue of their logical form alone. The argument isn't logically valid. The steps are truth-preserving (and necessarily so) because the conditional's antecedent is possible (and necessarily so), and this gives the misleading impression that the argument is **(p.280)** formally valid. But the antecedent's being possible is a metaphysical, not a logical fact. To make the argument formally valid, we need to add, as an additional premise, that the rocket's continuing on its course is possible. Then, given (SIC), the argument is formally valid.

12.5 *Reductio* Arguments

Another Williamsonian objection to non-vacuism about counterpossibles comes from *reductio* arguments (Williamson 2007, 2017). These are crucial to mathematics as it is practised. Williamson attempts to show that non-vacuists must hold current standard mathematical practice to be mistaken.

Although Williamson admits that *reductio* arguments need not be formulated in terms of counterfactuals, he takes it as legitimate to do so. And indeed, it is tempting to assert counterfactuals when reporting a particular line of reasoning by *reductio*: 'it can't be that A, because if it were that A, then it would be that B; but B is wrong, so A too must be'. That's valid on our semantics: $A \mathrel{\Box\!\!\rightarrow} B, \neg B \vDash \neg A$. (Weak Centring guarantees that the base world is an A-world if a B-world. But since we check for validity at a possible world, such a $\neg B$-world can't be an A-world, so must be a $\neg A$-world.)

The trouble stems from certain counterpossibles that can be used in *reductio* reasoning in this way. Since the reasoning is good, the counterpossibles ought to come out true. However, Williamson claims that non-vacuists cannot make good on this prediction. He considers the following examples (Williamson 2017):

(12.26) If there were a largest prime p, $p! + 1$ would be prime.

(12.27) If there were a largest prime p, $p! + 1$ would be composite.

(12.28) If there were a largest prime p, $p! + 1$ would be both prime and composite.

Williamson considers the following proof that there is no largest prime. First, establish (12.26) and (12.27) on their own merits, using **(p.281)** standard reasoning. Next, conclude (12.28) from them. Finally, appeal to our knowledge that no number is both prime and composite to conclude that there is no largest prime. As above, the final step of this reasoning is unproblematic for vacuists and non-vacuists alike. The alleged trouble for the non-vacuist is in getting (12.26)–(12.28) to come out true.

Williamson's worry is that a non-vacuist can't appeal to the usual mathematical reasoning we'd use to justify these. We could, ordinarily, reason that $p! + 1$ is not divisible by any $n \leq p$, so (if p is the largest prime) $p! + 1$ has no prime factors and must therefore be prime. However, we're assuming a number of mathematical results here. How can the non-vacuist be sure that they would hold, were there a largest prime? In general, non-vacuists deny that logical entailments carry across to valid counterfactuals: $A \vDash B$ does not imply $\vDash A \mathrel{\square\!\!\rightarrow} B$. So, if p were the largest prime, might it not be the case that $p! + 1$ *is* divisible by some $n \leq p$? How could we be sure either way? But if we don't have access to that reasoning, on the assumption that p is the largest prime, how can we ensure that (12.26) is true?

Similarly, we could ordinarily reason that, if p were the largest prime, then everything greater, including $p! + 1$, would be composite. We could then reason, on this basis, that $p! + 1$ would be both prime and composite. But what entitles the non-vacuist to this reasoning? Perhaps $p! + 1$ would not be greater than p, or perhaps Conjunction Introduction would not be valid, were there a greatest prime.

We think the answer to the puzzle lies in the context sensitivity of counterfactual utterances. Anna can truthfully say 'if I'd hit you, it would have hurt' (because she's got a mean punch); but she can also truthfully say 'if I'd hit you, it wouldn't have hurt' (because Anna wouldn't hurt anyone, so would have punched softly). Suppose, in a friendly conversation with no threat of violence, you ask Anna how strong she is. 'Let's just say', she replies, 'that if I'd hit you right then, it would have hurt'. That seems true (given how strong she is). But now suppose you and Anna are play-fighting. 'Watch it!', you say, as a mock punch comes a little close. 'Don't worry', she says, 'if I'd hit you then, it wouldn't have hurt'. In this way, 'the truth conditions **(p.282)** for counterfactuals … are a highly volatile matter, varying with every shift of context and interest' (Lewis 1973b, 92).

Any broadly Kratzer, Lewis, or Stalnaker-like approach to counterfactuals (Kratzer 1981, 1986, Stalnaker 1968, 1984) essentially involves two ingredients. It has an underlying space of worlds, plus some apparatus for focusing on the ones relevant to interpreting the counterfactual at hand in any particular case. *All* existing approaches to counterfactuals, vacuist and non-vacuist alike, take the second ingredient to be sensitive to the context in which a counterfactual occurs. There is simply no other way to get sensible results.

In the context of *reductio* reasoning, all the usual rules of reasoning must remain available. Similarly, if a counterfactual is uttered in that context, or in the context of reporting on such a proof, then the usual mathematical principles can be called upon to support the counterfactual. In the case of (12.26)–(12.28), in such contexts, conversational participants hold fixed what they know about the additive and multiplicative structure of the natural numbers. With such facts fixed, (12.26) and (12.27) follow easily, with (12.28) following by Conjunction Introduction.

But those mathematical facts need not be held fixed in every conversational context. We might be discussing mathematical finitism (as in Van Bendegem (1994)), and say, quite correctly, that if there had been a greatest number, there would have been a greatest prime number. In that context, we clearly are not retaining the mathematical fact that every number has a successor. Or we might be discussing what the physical world would be like if there were a largest prime number. Again, we cannot allow all of the facts of standard arithmetic to carry over.

As a consequence, valid logical and mathematical reasoning does not automatically carry over into counterfactual reasoning, as a matter of the logic of counterfactuals. A's entailing B does not imply that $A \mathrel{\square\!\!\rightarrow} B$ is valid. But, as we have seen, there may be contexts in which $A \mathrel{\square\!\!\rightarrow} B$ is true and justifiable on the basis of A's entailing B. They include the context of a *reductio* proof, or of explaining or reporting on such a proof. So the non-vacuist can **(p.283)** justify (12.26)–(12.28) and, more generally, she can justify this area of mathematical practice.

It may even be that vacuism about counterfactuals has the most trouble in capturing mathematical practice. The vacuist easily gets the result that (12.26)–(12.28) are true, since she takes all such counterpossibles to be trivially true. But now consider a mathematician explaining principles of constructive mathematics. In the following Q&A, someone raises an objection, 'but given what you're saying, had it been that $\neg\neg A$, then A, and so …'. 'No!' replies the speaker, pointing out that Double-Negation Elimination isn't constructively valid. She rejects as false the counterfactual, 'had constructive mathematics been correct, Double-Negation Elimination would still have been valid'. Her attitude seems to be part of accepted mathematical practice. Vacuists have trouble in

accommodating this. (They may take such counterfactuals to be unassertible, or otherwise out of place conversationally. But they can't capture the mathematician's attitude that those counterfactuals are false.)

12.6 Intuitions for Non-Vacuism

We've been defending non-vacuism about counterfactuals from a range of objections. But what are the positive arguments in its favour? Our support for non-vacuism largely rests on our ordinary-language judgements about the truth of a range of counterpossibles, such as the Hobbes-sentences (1.18) and (1.19), and the Anna-{Anna} sentences (12.1) and (12.2). (Jenny (2018) and Nolan (1997) offer further arguments.)

Williamson (2007, 2017) worries about this kind of motivation. He grants that the intuitions behind those judgementsd are present, but argues that they are not veridical. Here, we consider three Williamsonian arguments in this ballpark.

Thinking it Through

The first concerns the following example, due to Nolan (1997). (See also the discussion in Brogaard and Salerno (2013)). Suppose that **(p.284)** you were asked, 'what is 5 + 7?' and answer '12'. Now consider the following sentences:

> (12.29) If 5 + 7 were 13, you would have got that sum right.
>
> (12.30) If 5 + 7 were 13 you would have got that sum wrong.

(12.29) seems false and (12.30) true. But if (12.29) really is false then so is vacuism, since it's necessary that 5 + 7 isn't 13. Here is Williamson's response to this case:

> [Such examples] tend to fall apart when thought through. For example, if 5 + 7 were 13 then 5 + 6 would be 12, and so (by another eleven steps) 0 would be 1, so if the number of right answers I gave were 0, the number of right answers I gave would be 1. We prefer (12.30) to (12.29) because the argument for (12.30) is more obvious, but the argument for (12.29) is equally strong.
>
> (Williamson 2007, 172, our renumbering)

It seems to us, though, that the argument for (12.29) is not equally strong, for two reasons. First, having concluded that 0 = 1, it proceeds to substitute '1' for '0' within a counterfactual. But in general, that move is invalid (§12.4). Second, whether a particular chain of reasoning succeeds or fails in supporting the truth of a counterfactual depends on context, and in particular on which truths about the case need to be held fixed to legitimate the reasoning (§12.5).

In this case, all we need to hold fixed for (12.30) to be true is that the questioner asked what 5 + 7 is, that the answer given was 12, and that 12 is not 13. Williamson's argument for (12.29) needs to hold fixed all of that, plus facts about decrementing left and right addends. He must assume that 5 + 7 = 13 ⊢ 5 + 6 = 12 and its subtraction-generated cousins remain true, as well as facts connecting 'number of right answers' given to whether someone gets an answer right.

Some counterfactual contexts may support our retaining all of those facts, but not all will. The contexts in which (12.29) comes out true are thus a proper superset of those in which (12.30) comes out true. For to suppose that 5 + 7 is 13 is to suppose that the **(p.285)** additive structure of the numbers is something other than it actually is. Without some special context (such as during a mathematical proof, or reporting on mathematical *reductio* reasoning, as in §12.5), we have reason to expect that we should not hold fixed facts about incrementing and decrementing under such a supposition. So without some special context, we should expect that (12.30) is true whilst (12.29) is not. And it is no good for Williamson to place his argument within one of those special contexts. For as long as there is some context in which (12.29) is false, vacuism is too.

A Heuristic?

Let's go back to our first Hobbes-sentence:

> (1.18) If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would have cared.

No matter how we come at this sentence, we find it stubbornly wearing the appearance of a falsehood. Williamson's (2017) explanation for this appeals to a kind of error theory. We naturally take counterfactuals of the form $A \mathrel{\square\!\!\rightarrow} B$ and $A \mathrel{\square\!\!\rightarrow} \neg B$ to be contraries: 'if you were to win the lottery you would be happy' and 'if you were to win the lottery you would not be happy' cannot both be true. (Williamson suggests that we may confuse $A \mathrel{\square\!\!\rightarrow} \neg B$ with $\neg(A \mathrel{\square\!\!\rightarrow} B)$, thus taking them to be contradictories. But whether or not this is so, contrariety is all his explanation requires.)

This natural tendency is taken as the result of a fallible heuristic for counterfactual conditionals:

> (HCC*) If you accept one of $A \mathrel{\square\!\!\rightarrow} B$ and $A \mathrel{\square\!\!\rightarrow} \neg B$, reject the other.

(Williamson discusses two potential heuristics, (HCC) and (HCC*), with (HCC) telling us: If B and C are inconsistent, then treat $A \mathrel{\square\!\!\rightarrow} B$ and $A \mathrel{\square\!\!\rightarrow} C$ as inconsistent. Williamson prefers to use (HCC*) as it does not make use of the notion of inconsistency. We stick with Williamson's preference here.)

(p.286) If we evaluate $A \mathrel{\Box\!\!\rightarrow} \neg B$ to be true, then (HCC*) councils that we take $A \mathrel{\Box\!\!\rightarrow} B$ to be false. This is Williamson's explanation of why we take (1.18) to be false (erroneously, in his opinion).

It's unclear whether our reasoning in the case of (1.18) is guided by any such heuristic; and even if it were, it's doubtful that (HCC*) is the right heuristic. It has little plausibility when A is obviously impossible, as Williamson (2017, §6) acknowledges. It's easy to accept both of

(12.31) If it were raining and not raining, it would be raining.

(12.32) If it were raining and not raining, it would not be raining.

contrary to (HCC*)'s advice, for example.

Nevertheless, Williamson (2017, §6) maintains that (HCC*) plays a role when we evaluate a counterpossible to be false. We disagree. Suppose we're asked,

(12.33) If intuitionist logic were correct, would Excluded Middle be valid?

(Imagine we're feigning ignorance, for the benefit of our logic class.) We evaluate by considering situations in which intuitionist logic is correct. We know what these are like, because we understand the principles of intuitionist logic, BHK interpretations, Kripke semantics, and so on. In every such situation, Excluded Middle is not valid. So we judge that

(12.34) If intuitionist logic were correct, then Excluded Middle would not be valid.

is true. But in exactly the same way, we judge that

(12.35) If intuitionist logic were correct, then Excluded Middle would be valid.

(p.287) is false. The reasoning, via BHK interpretations or Kripke models or whatever, directly leads to our judgement of (12.35)'s falsity. It's not like we first have to work out what we think about (12.34), and then infer which stance to take on (12.35). Williamson's heuristic has nothing to do with it.

Exactly the same goes for Nolan's first Hobbes-sentence,

(1.18) If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would have cared.

According to Williamson, we assesses the conditional by imagining situations in which Hobbes squared the circle. In that situation, it's false that the sick South

American children care about Hobbes's achievement. But from there, we can infer, directly, that (1.18) is false. We needn't go via (1.19)'s truth and (HCC*).

Vacuous Quantification

A third argument against non-vacuist intuitions takes the form of an analogy between counterpossibles and vacuous universal quantification. The 'logically unsophisticated', according to Williamson (2007, 173), find it intuitive that 'every golden mountain is a valley' should be false, given that 'every golden mountain is a mountain' is true, on the grounds that *being a mountain* and *being a valley* are incompatible properties. However, both claims are true, vacuously, if there are no golden mountains. People extrapolate wrongly from familiar (non-vacuous) cases.

Williamson (2017, §6) expands the point. We know that dolphins don't have arms or legs and that unicorns have horns, so it's tempting to judge the following as false:

(12.36) Every dolphin in Oxford has arms and legs.

(12.37) Every unicorn is hornless.

Yet these claims are true on the standard treatment of quantifiers, because there are no unicorns or dolphins in Oxford.

(p.288) The intended analogy with vacuous quantification is clear. The 'logically unsophisticated', such as Franz and Mark, will intuitively judge counterpossibles like (1.18) and (12.2) to be false. But we make the same mistake as in the case of vacuous quantification. Since there are no situations which verify the antecedent, those counterfactuals are true. And, as in the quantification case, contrary pairs of counterfactuals will both be true when there are no situations to verify their antecedents.

But hang on a minute! We'd better not take the analogy too seriously. For by actualist lights, there are no situations verifying the antecedent of any counter-to-fact conditional. If a situation is contrary to fact, then it doesn't exist. If we understood conterfactuals as quantifiers over existing situations, then we'd end up treating them all as material conditionals, with all contrary-to-fact cases coming out trivially true. (One might insist, with Lewis, that there really do exist merely possible situations. But that extreme metaphysical view can't be required to make sense of counterfactuals.)

To make any sense, the analogy to quantification must be situated within a model (or a pretence, or whatever) in which there exist non-actual situations. Then we can take seriously the point that we make mistakes with vacuous quantification. Sure, if there are no situations to verify the antecedent, then we may have a true counterfactual which we're liable to judge as false. So the

question is, in general, are there (or should there be) such situations in our best semantic models? In the case of counterpossibles, the question becomes: should there be impossible situations in our semantic models? But this is a key point at issue in the vacuist-non-vacuist debate. It seems dialectically illegitimate to assume that only possible situations may play a role in our models.

Chapter Summary

There are prima facie reasons to think that *vacuism*, the view that all counterpossibles are trivially true, is incorrect (§12.1). We then **(p.289)** offered an impossible worlds semantics for counterfactuals, which makes room for non-trivial counterpossibles (§12.2). The semantics raises a number of questions. One principle which pins down its application is the *Strangeness of Impossibility* condition, which says that, for any given possible world, any impossible worlds is further away from it than any possible world is (§12.3). We discussed a number of Williamson's objections to the non-vacuist approach in the context of (SIC), and argued that they can be overcome.

We then raised the question of whether counterfactuals in general (including counterpossibles) should permit the substitution of rigidly coreferential terms, again by considering Williamson's arguments against non-vacuism (§12.4). The third case that Williamson makes against non-vacuism is that it does not make good sense of the way *reductio* arguments are used in mathematical practice. We showed how non-vacuists can resist this argument (§12.5). Having defended non-vacusim against Williamson's objections, we then considered a range of arguments in its favour (§12.6). **(p.290)**