Logical Theory-Choice: the Case of Vacuous Counterfactuals

Graham Priest
Philosophy Programs
CUNY Graduate Center and the University of Melbourne

1 Introduction

Logic—in one important sense of that polysemic word—is a theory about what follows from what. It is a theory about what inferences are valid and why. Such theories have been advanced in Western logic for over two and a half thousand years, the theories disagreeing with each other in numerous different ways.[1] We face, then, the question of how one should choose the best theory: what is the procedure for rational theory-choice? Elsewhere, I have advanced an answer to the question, and argued for it.[2] Theory-choice in logic is just a special case of theory-choice in general. The precise details of the implementation may differ, depending on the area of theory-choice in question (science, metaphysics, ethics, aesthetics, logic); but there is a uniform and general framework for theory-choice. The main purpose of this paper is to illustrate this with one particular case-study.

This is as follows. There is at present a certain dispute about counterfactuals taking place. What is at issue is whether counterfactuals with necessarily false antecedents are all true.[3] Thus, for example, consider the counterfactuals:

[1]The matter is discussed in detail in Priest (2014).

[2]See Priest (2016a).

[3]I shall talk of counterfactuals. However, I take the distinction between indicative and subjunctive conditionals to be a spurious one. (See Priest (2018).) So I might just as well have talked of conditionals, *simpliciter*. However, this is not the place to go into that matter.

- If intuitionist logic were correct, the principle of explosion ($A, \neg A \models B$) would not be valid.
- If Hobbes had squared the circle, all sick children in the Andes would have cared.

The antecedents are necessarily false—or so we may assume. Some hold that such counterfactuals are vacuously true, appearances notwithstanding. Let us call such people *vacuists*. Others hold that some counterfactuals with necessarily false antecedents are true; some are false: it just depends on their contents. Let us call such people *non-vacuists*. As a notable representative of the vacuists, I will take Tim Williamson and the case he makes in (2007) (esp. ch. 5) and (2017). On the other side, I will take the position defended in Berto, French, Priest, and Ripley (2018). I will argue (unsurprisingly) that the better choice is Non-Vacuism. That, however, is a subsidiary aim of this paper. The main point is to illustrate the method of theory-choice at issue.

The paper falls into two main parts. The first sets out the preliminary details necessary to understand the case-study. Specifically, I will explain the method of theory-choice to be deployed; I will then explain the two theories of counterfactuals to which it is to be applied. Armed with this background, the second part of the paper provides the application of the method to the case at issue.

2 Preliminaries

2.1 Logical Theory-Choice

So let us start with theory-choice. Theories are proposed to explain something or other, that is, to account for some data. In the case of logic, the data is the data we have about some inferences which look valid and some others which do not. Thus, the inference:

- You are in Rome.
- If you are in Rome, you are in Italy.
- So you are in Italy.

appears valid. While the inference:

- You are in Italy.
- If you are in Rome you are in Italy.
- So you are in Rome.

does not.

However, most theories do not explain all the data, which are, in any case fallible. Moreover, it can happen that different theories explain the same data. Adequacy to the data can therefore be only one relevant criterion—even if the most important. There are other criteria standardly applied in theory choice. Discussion of the criteria and their articulation are standard fare in the philosophy of science.[4] For the sake of definiteness here, let us take the criteria to be as follows:

- adequacy to the data
- consistency
- simplicity
- power
- unifying power

Now, the criteria will not, in general, all pull in the same direction. Thus, one theory, T_1, may do better justice to the data than another, T_2. Yet T_2 may be much simpler than T_1. We have to aggregate the performance of different theories on the various criteria in some way, bearing in mind that the criteria may be of different degrees of importance. The best theory, if there is one, will be the one which comes out best overall.

2.2 A Formal Model

We can provide a simple (perhaps better, simplistic) model of the aggregation involved as follows. Let the criteria in question be $c_1, ..., c_n$. The measuring scale is, to a large extent, a matter of convention, but for the sake of determinacy, let this be the integers between +10 and -10 (+10 being the best). For every criterion, c, there is a measure function, μ_c, which maps every theory

[4]See, e.g., Quine and Ullian (1978), Lycan (1988).

on the table, T, to a value in the scale. So $\mu_c(T) \in [-10, +10]$. Further, every criterion, c, will have a weight of importance, w_c. We can take weights to be measured on the (also conventional) scale $[+1, +10]$; so $w_c \in [+1, +10]$.

We may now define the *rationality index* of theory T, $\rho(T)$, as follows:

$$\rho(T) = w_{c_1}\mu_{c_1}(T) + ... + w_{c_n}\mu_{c_n}(T)$$

If the theories on the table are T_1, ..., T_k, the rationally preferable one is that with the highest rationality index. If there is a tie, we may suspend judgment, or just choose at random.[5]

2.3 Vacuism

So much for the method of logical theory-choice. I will now explain the two theories of counterfactuals to which we will apply the method, starting with the vacuist theory. This is a standard theory of counterfactuals for a propositional language, which deploys possible worlds.[6]

Our language has the connectives $\wedge, \vee, \neg, >, \Box, \Diamond$, with their usual syntax. ($>$ is the counterfactual conditional.) Let Π be the set of propositional parameters, and Φ be the set of formulas.

An interpretation is a structure $\langle P, \{R_A : A \in \Phi\}, \nu\rangle$, where:

- P is the set of (possible) worlds
- for every formula, $A \in \Phi$, R_A is a binary relation on P
- $w_1 R_A w_2$ means that w_2 is a world *ceteris paribus* like w_1, except that A is true[7]
- for every $p \in \Pi$ and $w \in P$, $\nu_w(p) = 1$ or $\nu_w(p) = 0$

Given an interpretation, truth at a world ($\Vdash$) is defined recursively as follows:

- $w \Vdash p$ iff $\nu_w(p) = 1$

[5]We may assume, in the case to be pursued, that the metalogic, that is, the underlying logic of the computation, is classical. What happens in more general cases, where one is trying to adjudicate between theories which have different metalogics? The matter is taken up in Priest (2016a), Section 3.4.

[6]See, for example, Priest (2008), ch. 5.

[7]The *ceteris paribus* notion is certainly context-sensitive. However, this context-sensitivity plays no role in the formal semantics.

- $w \Vdash \neg A$ iff $w \nVdash A$
- $w \Vdash A \wedge B$ iff $w \Vdash A$ and $w \Vdash B$
- $w \Vdash A \vee B$ iff $w \Vdash A$ or $w \Vdash B$
- $w \Vdash \Box A$ iff for all $w' \in P$, $w' \Vdash A$
- $w \Vdash \Diamond A$ iff for some $w' \in P$, $w' \Vdash A$
- $w \Vdash A > B$ iff for all w' such that wR_Aw', $w' \Vdash B$

Validity is defined in the standard way:

- $\Sigma \models A$ iff for every interpretation, and for every $w \in P$: if $w \Vdash B$ for all $B \in \Sigma$, then $w \Vdash A$

The logic of the modal operators is $S5$. As things stand, there are no constraints on the various R_As. However, the following would appear to be mandated by the meaning of the accessibility relation.

If R_A takes us to a world, A is true there:

- if wR_Aw' then $w' \Vdash A$

This constraint verifies the inference:

- $\models A > A$

Next, if A is true at w, then w is one of the worlds that is *ceteris paribus* the same as w except that A is true there:

- If $w \Vdash A$ then wR_Aw

This verifies the inference:

- $A, A > B \models B$

Whether one should demand other constraints on the R_As, is a matter we need not go into go here.[8] The important point is that the following inference is clearly valid:

- $\neg\Diamond A \models A > B$

This is the vacuist thesis.

[8]Similarity-sphere semantics for counterfactuals, more familiar to most people than the above semantics, can be obtained by the addition of further constraints. (See Priest (2008), ch. 5.) The notion of similarity is, as is well known, just as much context-dependent as that of being *ceteris paribus* the same.

2.4 Non-Vacuism

Let us now turn to the non-vacuist theory. This is the same, except that the semantics augments the collection of possible worlds with a collection of impossible worlds, I.[9] At possible worlds, the truth conditions are the same; but at impossible worlds, truth values are assigned non-recursively. That is, ν is extended by the condition:

- for every $A \in \Phi$ and $w \in I$, $\nu_w(A) = 1$ or $\nu_w(A) = 0$

and if $w \in I$:

- $w \Vdash A$ iff $\nu_w(A) = 1$

Note that each R_A is now a binary relation on $W = P \cup I$. So a possible world may access an impossible world under R_A—which is exactly what one would expect if A is logically impossible. Note, also, that validity is still defined as truth preservation at all *possible* worlds of all interpretations.

One may expect each of the R_As to satisfy the two conditions mentioned before, and for exactly the same reason. We should also expect them to satisfy the following condition, the *Integrity of the Possible*, IP:

- if $x \Vdash A$ for some $x \in P$ then: if $w \in P$ and wR_Aw', $w' \in P$

If w is a possible world, and A is a possible condition, a world that is *ceteris paribus* the same as w except that A holds, is itself possible. As is easy to check, IP verifies the inference:

- $\Diamond A, A > B \models \Diamond B$

Finally, as is easy to see, under these semantics:

- $\neg\Diamond p \not\models p > q$

Just take an interpretation where p is false at every possible world, but where $w \in P$, $w' \in I$, wR_pw', p is true at w', but q is not. Thus, we have Non-Vacuism.

[9]For further discussion of the notion of impossible world, see Priest (2016b), which appeared with minor modifications as Priest (2016c), ch. 9.

3 Theory Comparison

We may now turn to applying our account of theory choice to these two theories at hand. The following arguments are abstracted from the references cited in Section 1. I focus here on the most important points. In order not to loose the wood for the trees, I do not attempt to follow the arguments through every twist and turn of their dialectics.

We need to evaluate how our two theories fare on the various criteria in play. The most complex of these is adequacy to the data. The other criteria are relatively straightforward, so let me deal with these first.

3.1 Other Criteria

The first of these is the easiest: *Consistency.* Both theories are consistent, so there is nothing to choose between them.

The next criterion is *Simplicity.* We might distinguish here between *Conceptual Simplicity* and *Ontological Simplicity*. *Conceptual Simplicity*: Clearly, the machinery of impossible worlds adds to the complexity of the possible world semantics. The addition, though, is itself of a very simple kind. So Vacuism is simpler conceptually, but only sightly. *Ontological Simplicity*: Non-Vacuism invokes the possible worlds of Vacuism, but, in addition, impossible worlds. Depending on how one understands worlds, these could be an extra kind of entity. Thus, some theories of worlds take possible and impossible worlds to be different kinds of things; some take them to be the same kind of thing.[10] Personally, I think that they are the same. However, this is not the place to go into that subject here. So let me (in the spirit of magnanimity!) give the decision on that matter to the vacuist. Vacuism, then, is ontologically simpler. A caveat should be noted, though. Ockham's Razor says that entities should not be multiplied *beyond necessity.* If it turns out that the other criteria speak to the necessity of impossible worlds, the weight of this criterion is diminshed. At any rate, in both senses of simplicity, Vacuism is simpler, though perhaps not overwhelmingly so.

The next criterion is *Power. Prima facie*, this criterion speaks strongly in favour of Vacuism. The vacuist interpretations are a subset of the non-vacuist ones (those where $I = \emptyset$). The inferences validated by the non-vacuist semantics are therefore a subset of those validated by the vacuist ones. Indeed, they

[10] For a discsussion of the matter, see Berto (2013).

are a very proper subset. The impossible-world semantics invalidates even very simple inferences of standard counterfactual logic. Thus, for example:

- $\not\models (A \wedge B) > A$
- $\not\models A > (A \vee B)$

An impossible antecedent may take us to an impossible world where anything is, erm, possible.

However, matters are not so straightforward. As is easy to see, given IP, these inferences can be regained by adding the premise that the antecedents are possible.

- $\Diamond(A \wedge B) \models (A \wedge B) > A$
- $\Diamond A \models A > (A \wedge B)$

Indeed, any inference that is valid in the possible-world semantics can be recaptured by adding the premises that the antecedents of all the conditionals involved are possible. IP then assures us that in evaluating formulas, we never leave the domain of possible worlds. In other words, business is as normal, and no power is lost. The two theories perform equally on their inferential power.

There is a possible reply at this point. Yes, IP does deliver the power of the possible world semantics. But IP is *ad hoc*. So the recapture comes at a methodological cost. (*Ad hoc*ness is a certain kind of failure of simplicity.) There is a clear counter-reply, though. IP is not at all *ad hoc*. It is mandated by the very understanding of the accessibility relation. As already noted, if w is a possible world, and A is a possible condition, then one should expect a world that is *ceteris paribus* the same as w, except that A holds, to be a possible.

The upshot of these considerations is that the two semantics perform about the same on matters of inferential power.

The final (other) criterion is *Unifying Power*. Impossible worlds may add a complexity to the semantics, but they also have applications in quite distinct areas.[11] For example, they are required to give a satisfactory worlds-account of intentional operators, such as *believe that*, or *desire that*. Otherwise, it transpires that one believes and desires everything that is necessarily

[11]On the use of impossible worlds in a variety of areas, see Priest (1997), and the papers in that issue of the *Notre Dame Journal of Formal Logic*.

true.[12] Similarly, impossible worlds are required to give a worlds-account of the content of statements. Otherwise it transpires that all necessary truths have the same content, as do all necessary falsehoods.

One might simply grit one's teeth and accept these odd conclusions. But this is a manifestly *ad hoc* move. Alternatively, one might try to avoid these consequences in a worlds-account of intentionality and content, by complicating the machinery in some other way.[13] But this also comes with a methodological cost: namely much additional complexity.

In sum, then, the impossible-world semantics has a unifying power, bringing together counterfactuals, intentionality, content, in a way that a merely possible-world semantics cannot match. On this criterion, then, Non-Vacuism is much the preferable.

Before we move on, let us take stock. I summarise the conclusions reached thus far in a table. (A plus denotes an advantage; a blank denotes none.) I have assigned the criteria rough weights. There is certainly room for some discussion here; but I think that most people would find them roughly right.

	Vacuism	*Non-Vacuism*	*Weight*
Consistency			*high*
Simplicity	+ (Slightly)		*lowish*
Power			*medium*
Unifying Power		+	*medium*
Adequacy to Data	?	?	*very high*

3.2 Adequacy to the Data 1

So far, then, there are no really decisive considerations. So let us turn to the criterion of *Adequacy to the Data*. As the weightings in the above table indicate, this is the most important criterion. The whole *purpose* of a theory is, after all, to account for the relevant data.

Prima facie, the situation here is very clear, and speaks strongly in favour of Non-Vacuism. Thus, consider the pairs:

[**1**] If intuitionist logic is correct, the Principle of Excluded Middle (PEM) is invalid.

[12] See Priest (2016c).
[13] For example, by moving to structured meanings, as in Cresswell (1985).

[**2**] If intuitionist logic is correct, Explosion is invalid.

[**3**] If you were to prove [refute] Goldbach's conjecture, you would become a famous mathematician

[**4**] If you were to prove [refute] Goldbach's conjecture, I would give you my life's savings

The first of each pair seems clearly true, and the second, clearly false.[14] Given these examples, and many others like them, Vacuism seems highly inadequate to the data.

There is a possible reply, however; that is to challenge the data—as does Williamson. Of course, simply to deny the data is thoroughly *ad hoc*. But the *ad hoc*ness can be removed if one can give an independent explanation of why we are mistaken about the data. Williamson's explanation is that when we take a counterfactual with an impossible antecedent, $A > B$, to be false, it is because we evaluate $A > \neg B$ as true first and then apply the heuristic:

- if $A > \neg B$ is true, $A > B$ is false

The reply itself is inadequate, however; and this is so for at least three reasons.

First, it presupposes that we evaluate a particular one of the conditionals first, but there is no reason why we should not start by evaluating the other. Thus, we could just as well have evaluated $A > \neg B$ first, found it true, and then concluded that $A > B$ is false. We would then have ended up with the opposite conclusion.

Secondly, if one were to apply Williamson's heuristic, it could not be the case that counterfactuals of the form $A > B$ and $A > \neg B$ appear both to be true or both to be false; but there are examples of such. For truth:

- If it were and were not raining, it would be raining.

- If it were and were not raining, it would not be raining.

For falsity:

- If it were raining and not raining, it would be Tuesday.

[14]In the Goldbach case, the antecedent is whichever statement is impossible; and the warrant for the second counterfactual is my *say-so*!

- If it were raining and not raining, it would not be Tuesday.

Thirdly, and most importantly, the heuristic is *not* normally deployed in the evaluation of an apparently false counterfactual. Both true and false counterfactuals are evaluated *in exactly the same way*: directly. We imagine the situation where the antecedent is true, and see if the consequent is true there. Thus, to evaluate [1] we consider a world where intuitionist logic holds. We know what this is like, since BHK semantics, Kripke models, etc, are well understood. And we know that in such worlds, the PEM fails. Similarly, in evaluating [2], we consider a world in which intuitionist logic holds, and we know that Explosion is still valid there.

As an interim judgment, then, the preference for Non-Vacuism on this criterion still stands.

3.3 Adequacy to the Data 2

But that is not an end of the matter. One might reply (as does Williamson) that there is other data that the non-vacuist cannot explain, whereas Vacuism does. Matters, then, can be seen as more equal.[15]

One might proffer a couple of different examples of data that the non-vacuist cannot account for. Here is the first. In impossible-world semantics, we have the following:[16]

- $a = b, A > Pa \not\models A > Pb$

But we use inferences like this all the time in our reasoning. For example, we may reason as follows:

- If the rocket had continued on its course, it would have hit Hesperus.
- Hesperus is Phosphorus.
- If the rocket had continued on its course, it would have hit Phosphorus.

[15] This might open the issue of which data are the more important; but we need not go into this here.

[16] This is not obvious from the semantics sketched in 2.4, since no semantics for identity is specified. However, when this is done, the claim is correct. The central point is that if A is impossible, then evaluating the conditionals takes us to impossible worlds; and there is no reason why Pa and Pb must stand or fall together, even though a actually is b. We need not go into the details here.

The non-vacuist cannot account for this.

The reply to the argument is obvious, however. We have already seen it at work in Section 3.1. IP assures us that the formal inference *is* valid if we add the extra premise that $\Diamond A$—which is obviously true in the example given. The additional premise ensures that we do not venture into impossible worlds; and at possible worlds, identity behaves normally.

Moreover, the failure of the Substitutivity of Identicals for counterfactuals with impossible antecedents in not at all *ad hoc*. It is what one should *expect* to happen. Merely consider:

- If Hesperus were not Phosphorus, modern physics would be badly mistaken.
- Hesperus is Phosphorus.
- So if Hesperus were not Hesperus, modern physics would be badly mistaken.

The conclusion is wrong. If Hesperus were not Hesperus, it would not be modern physics that is mistaken: it would be modern logic.

A quite different putative example of a piece of reasoning that the non-vacuist cannot account for concerns arguments by *reductio ad absurdum*. For example, in the course of proving that there is an infinitude of prime numbers, one might invoke the conditionals:

- If p were the largest prime number, $p! + 1$ would be prime.
- If p were the largest prime number, $p! + 1$ would not be prime.

These are, of course, true for the vacuist; but not for a non-vacuist—or so it is claimed.

The first reply is that the non-vacuist does not have to take these conditionals to be true: they are merely *façons de parler*. Thus, one might take the first of these to express the following:

- Let p be the greatest prime. Then $p! + 1$ is prime.

This is simply a statement of deducibility, and perfectly acceptable to a non-vacuist.

The second reply is that a non-vacuist *can* account for the truth of these counterfactuals anyway. In worlds that are *ceteris paribus* the same as ours except that p is the greatest prime, the basic facts about multiplication, etc., still hold. So then the consequent *is* true.

The case concerning adequacy to the data on the side of the vacuist therefore breaks down, leaving just the case on the side of the non-vacuist. This criterion therefore speaks heavily in favour of Non-Vacuism.

3.4 Summary

Let us bring all the considerations of Section 3 together, in the form of a completed table.

	Vacuism	*Non-Vacuism*	*Weight*
Consistency			*high*
Simplicity	+ (Slightly)		*lowish*
Power			*medium*
Unifying Power		+	*medium*
Adequacy to Data		+	*very high*

As should be clear, even without exact details for the weights, ρ(Non-Vaculism)> ρ(Vacuism). Hence, Non-Vacuism is the better theory.

4 Conclusion

That conclusion is, of course, interesting and important. However, to establish it was not the main aim of the paper. The main aim, as stated right at the start, was to illustrate the methodology of theory choice outlined in Sections 2.1 and 2.2.

Of course, when logicians argue for and against different theories, they do not usually do the explicit cost-benefit analysis required by the methodology. They engage in the dialectic back-and-forth of arguments of the kind we have seen. Naturally, such things are important. However, we can now see that they are important because they fit into the standard methodological framework of theory-choice. Perhaps logicians do not have a fully articulated understanding of the matter; perhaps the understanding is an inchoate one. Be that as it may, we see that the framework makes sense of the considerations that parties in a debate about choice of logic *do* put forward. The present case study not only, therefore, illustrates the methodology in question, but also speaks in its favour, by explaining this central piece of data.[17]

References

[1] Berto, F. (2013), 'Impossible Worlds', in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/impossible-worlds/.

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

[3] Cresswell, M. J. (1985), *Structured Meanings*, Cambridge, MA: MIT Press.

[4] Lycan, W. (1988), *Judgement and Justification*, Cambridge: Cambridge University Press.

[5] Priest, G. (1997), 'Editor's Introduction', *Notre Dame Journal of Formal Logic* 38: 481-87.

[6] Priest, G. (2008), *Introduction to Non-Classical Logic: From If to Is*, Cambridge: Cambridge University Press.

[7] Priest, G. (2014), 'Revising Logic', ch. 12 of P. Rush (ed.), *The Metaphysics of Logic*, Cambridge: Cambridge University Press.

[8] Priest, G. (2016a), 'Logical Disputes and the *a Priori*', *Logique et Analyse* 236: 347-66; and *Principios: Rivista de Filosofia* 23: 29-57.

[9] Priest, G. (2016b) 'Thinking the Impossible', *Philosophical Studies*, 173: 2649-2662.

[10] Priest, G. (2016c), *Towards Non-Being*, 2nd edn., Oxford: Oxford University Press.

[17] Versions of this paper were read at the University of Bergen in November 2016, and the University of Edinburgh, April 2017. I am grateful to members of the audiences there for their helpful comments and suggestions. Many thanks also go to an anonymous referee for this journal.

[11] Priest, G. (2018), 'Some New Thoughts on Conditionals', *Topoi* 27:369-377.

[12] Quine, W., and Ullian, J. (1978), *The Web of Belief*, 2nd edn., New York, NY: McGraw-Hill.

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

[14] Williamson, T. (2017), 'Counterpossibles in Semantics and Metaphysics', *Argumenta* 4 (online link).