

MUST HYPOTHETICAL COUNTEREXAMPLES BE POSSIBLE?

Hypothetical counterexamples are a mainstay in philosophical debate, perhaps the most common tactic in attempting to establish some philosophical position, across a wide range of subdisciplines. When tailoring these counterexamples, philosophers choose possible but usually non-actual situations. No one has seriously considered whether these counterexamples must actually be metaphysically possible situations, however. This paper argues that there is no theoretical reason to insist that counterexamples be possible, and speculates about some implications of this conclusion for various debates within philosophy, briefly that this discovery will be a boon to particularists in normative areas of philosophy.

I. THE NATURE OF HYPOTHETICAL COUNTEREXAMPLES

It is easy for seasoned philosophers to recall a long list of hypothetical counterexamples in the philosophical literature. In ethics, there are murderers at the door and Southern sheriffs.¹ In epistemology, there are Gettier cases and reliable clairvoyants.² In metaphysics, there are malfunctioning tele-transporters, people who fail to realize that a door is locked, and lumps and statues.³ Philosophers probably use hypothetical (rather than actual) counterexamples because there is no need to seek out and identify real-world objects, especially when such objects might be physically impossible with current technology, such as a tele-transporter. The other salient feature of

¹ The “murderer” objection may be due originally to Benjamin Constant. See I. Kant (tr. J. Ellington), *Grounding for the Metaphysics of Morals*, New York, NY: Hackett, [1785] 1993: 63-7. See also H. J. McCloskey, “An Examination of Restricted Utilitarianism,” *Philosophical Review* 66 (1957), 466-485.

² E. Gettier, “Is Justified True Belief Knowledge?,” *Analysis* 23 (1963), 121-3; L. BonJour, *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press, 1985: Chapter 5.

³ Compare D. Parfit, *Reasons and Persons*, Oxford: Oxford University Press, 1984. See also H. G. Frankfurt, “Alternate Possibilities and Moral Responsibility,” *The Journal of Philosophy* 66 (1969), 829-39, on locked rooms and M. REa (ed.), *Material Constitution: A Reader*, Lanham, MD: Rowman & Littlefield, 1997.

these counterexamples is that they tend to be, as far as anyone knows, metaphysically possible.⁴ Even if tele-transporters are impossible given current technology—and may even be impossible given the physical laws of the universe⁵—there is no in-principle problem with imagining them, and to state that they exist does not appear to violate any theorems of standard logic. As far as anyone knows, there is a sense in which they ‘could’ exist. But it is not so easy to speculate about *why* philosophers choose possible (rather than impossible) counterexamples. Maybe they are simply easier to imagine. If that is the only reason, then there is no in-principle difficulty with employing impossible counterexamples, as long as philosophers find themselves concluding at least something relevant about the imagined situation.

This paper argues for the revisionary conclusion that there is, after all, no particular reason to insist upon the *Possibility Condition*:

A hypothetical counterexample must be possible in order to be probative
And indeed, there is some reason to reject it. The most straightforward terminology risks a bit of confusion by borrowing a term already in use elsewhere: for this paper, ‘possibilists’ believe hypothetical counterexamples must be possible in order to be probative, and ‘apossibilists’ deny this requirement.⁶ To take a very simple example: Suppose one were to hold the outlandish theory that nothing can be imagined to be circular. For possibilists, the only way to object to this theory would be to point to possible or actual objects that are circular. A possibilist would, upon presented with this example, abandon the thesis that nothing is circular. In contrast, an apossibilist would even abandon the theory upon recognizing the hypothetical (impossible) counterexample of a square

⁴ Possibility is vital for the discussion in T. Williamson, *The Philosophy of Philosophy*, Malden, MA: Blackwell, 2007. Compare philosophical zombies and the debate over their possibility; see D. J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, New York and Oxford: Oxford University Press, 1996.

⁵ Perhaps given Heisenbergian uncertainty, for example.

⁶ Compare R. M. Adams, “Actualism and Thisness,” *Synthese* 49 (1981): 3–41.

circle. A possibilist would insist that a square circle does not provide a counterexample to the theory that nothing can be imagined to be circular.

It is helpful before proceeding further to be explicit about what a hypothetical counterexample actually is. For this paper's purposes, a hypothetical counterexample takes the following form, the *Counterexample Schema*:

1. If theory *T* is true, then if *C* were to obtain, then *P* would obtain.⁷
2. But it is false that if *C* were to obtain, then *P* would obtain.
3. Therefore, *T* is false.

Here, *C* is some condition or situation, such as a murderer at the door or a malfunctioning tele-transporter. The point of a hypothetical counterexample, then, is to charge a theory with making a falsified prediction; some theory predicts one outcome or conclusion, but (perhaps intuitively) a contrary outcome or conclusion is what would actually obtain. And the pertinent question for this paper is whether counterexamples work when the value of *C* is something metaphysically impossible.

So far, of course, these are rather abstract remarks with little in the way of real-life philosophical "bite." It may be helpful to remain at this level of abstraction for now in order to establish a possibilism. After the arguments for a possibilism are presented, this paper will explore further implications for actual philosophical debates, which will help to put flesh on this skeleton of a position. It turns out that these implications are, in some cases, quite revisionary indeed.

II. FOR AND AGAINST THE POSSIBILITY CONDITION

Possibilists impose the Possibility Condition, according to which hypothetical counterexamples must be possible in order to be probative. Apossibilists do not impose this condition. As stated earlier,

⁷ Of course, the conditional is subjunctive because the counterexample is hypothetical, not actual.

most philosophers seem to gravitate toward possibilism without putting much thought into the question. After all, why bother with impossible counterexamples when the possible ones seem to be working just fine? In general, one might expect the abandonment of the Possibility Condition to coincide with dialectical gains for philosophical skeptics and particularists of various stripes, since now general theories will be open to many more counterexamples.⁸ So this is what is at stake: possibilism will prove amenable to theory-builders, and apossibilism will prove amenable to theory-breakers. These expectations will not be taken as reasons in their own right to prefer possibilism or apossibilism, since that would require an independent defense of one type of approach. But the expectations will be borne out in the final section of this paper.

How, then, would one adjudicate this debate? A straightforward approach would be to hold that the most basic position makes no claim at all about whether the Possibility Condition is required, and argue that anyone who wishes to take any particular position here bears the burden of proof. Someone taking this approach could cheerfully offer a host of impossible counterexamples, and then sit back and wait for the possibilist to argue against those counterexamples, whether she chooses to stake her position on their impossibility. After all, few philosophers who offer counterexamples in general take the time to justify the method in general, and so apossibilists might well continue with this tradition: 'I'm going to offer a counterexample, and not explain why you should accept counterexamples, since no one else tries to explain this either.' Then the possibilist can attempt to offer specific objections, some of which may be based upon the impossibility of the situation identified in the hypothetical counterexample.

This paper will proceed by attempting to answer arguments against impossible counterexamples, and then offering some positive remarks about why one cannot oppose impossible counterexamples without opposing the method of counterexample as a whole. There

⁸ For example, generalists in ethics say that ethics comprises moral principles; particularists deny this. Compare J. Dancy, *Ethics Without Principles*, Oxford, UK: Oxford University Press, 2004.

may be serious problems with the method of counterexample in general, but this paper is not intended to answer them. Instead, the aim is simply to show that there are no special difficulties for impossible counterexamples, at least in many useful applications of such counterexamples.⁹

1. *What's Not Wrong With Impossible Counterexamples*

Perhaps the simplest way to object to impossible counterexamples would be to deny that anyone knows how to evaluate subjunctive conditionals with metaphysically possible antecedents. Then the Counterexample Schema would be invalid when $C =$ some impossible condition or situation. But the apossibilist can respond that there actually are ways of evaluating subjunctive conditionals with impossible antecedents, indeed with non-trivially true results. After all, one worry would be that if a subjunctive conditional were treated as a mere necessary conditional, it would come out trivially true. So it is important, first of all, to note that subjunctive conditionals can only come out non-trivially true if they are treated as something *sui generis*, rather than just necessary conditionals. If they are, then one can begin to suggest ways to evaluate them.

The first step is to note that that it is a theorem of all standard predicate logics that:

$$(\forall\Phi)(\forall\Psi)(\forall\varphi)[(\Phi\varphi \wedge \Psi\varphi) \supset \Phi\varphi].$$

Call this the ‘ \wedge -elimination theorem.’ In English, the theorem states that if some object satisfies two predicates, it follows that it satisfies one of those predicates. Since everyone accepts this already, this may give one a relatively easy way of evaluating some conditionals with impossible antecedents. After all, it follows straightforwardly that:

If x is square and circular, then x is square.

⁹ For what might be more general problems with the method of counterexample, compare (for example) S. Kagan, “The Additive Fallacy,” *Ethics* 99 (1988): 5-31.

Apparently everyone is already committed to the \wedge -elimination theorem, and therefore everyone should be committed to any substitution-instance of it. One might of course deny that a logic containing that theorem is sound if it is intended to apply to models in which Φ and Ψ are uncoinstantiable. Perhaps there is an argument for that conclusion, but it is so far unknown. So at first glance, even widely accepted languages of logic provide a way to evaluate some subjunctive conditionals with impossible antecedents.

There are also intuitive reasons to think one can evaluate subjunctive conditionals with impossible antecedents. Specifically, some of these conditionals can seem non-trivially true or false. After all, it might just seem true that if there were a square circle, there would be a shape, or if $2+2$ were to equal 5, then accurate calculators would report that $2+2=5$. The onus is on the possibilist to explain why, specifically, one should not simply evaluate these as *true*, and non-trivially so. What is an accurate calculator, other than one that correctly reports mathematical equalities?

What is more, real-world philosophers already accept as non-trivially true some subjunctive conditionals with impossible antecedents. For example, Anselmian theists believe that God is a metaphysically necessary being. They also believe God caused the universe to exist. Now, one analysis of causation is counterfactual, from which it would follow that if God had not existed, the universe would not have existed. (Which theists would ever deny this?) But of course this is just an assertion of the non-trivial truth of a subjunctive conditional with an antecedent that Anselmians say is metaphysically impossible. Another example occurs in normative ethics. Presumably many normative ethicists think their ethical theories are necessarily true; one does not find, for example, normative ethicists commonly asserting that deontology is true, but consequentialism might have been true. But deontologists would presumably say that if utilitarianism had been true, it would have been obligatory to maximize utility. This is not false; it is not even merely trivially true. It is just true.

So philosophers commonly think that it is possible to evaluate subjunctive conditionals with impossible antecedents, and there is so far no reason to disagree with them.

Now, the aforementioned cases are apparently examples wherein the antecedent in question is taken to be metaphysically impossible but logically possible. There is no purely *logical* contradiction in asserting that the Anselmian God does not exist or that deontology is false. That is, no standard languages of logic have Anselmian theism as a theorem. On the other hand, it may be quite a different story to attempt to evaluate conditionals with explicit logical contradictions in their antecedents. If there were a square non-square, then—what? Is there any consequent to be non-trivially concluded? Well, perhaps that something square would exist. This example is likely to be more controversial, however, so when attempting to do philosophy without the Possibility Requirement, perhaps one will have to employ only the impossible counterexamples, the antecedents of which are nevertheless logically possible. This paper will return to this point in the third section.

Another possible difficulty with employing impossible counterexamples would be that the philosopher whose theory is a target of the counterexample might charge the apossibilist with question-begging. After all, to take another simple case, suppose one's theory is that there cannot be a square circle, and the apossibilist uses the example of a square circle as a counterexample. Surely this should not be persuasive; it is simply a case of denying the very theory at issue. And after all, the person who denies the possibility of square circles can insist that her theory is unscathed, since the apossibilist admits that square circles are impossible. So perhaps impossible counterexamples cannot be used in this simplistic way; instead, their use is less direct. Consider, then, a theory according to which one could never be justified in believing in square circles. Perhaps in this case, one might respond that if one were to perceive a square circle, one would thereby be justified in believing in it. This does not seem to presuppose the falsity of the theory at issue after all. It shows that there is

nothing in principle wrong with the idea that someone could be justified in believing in square circles, even if in practice this would probably never occur.

The overall lessons so far are three: First, there is no purely logical bar to employing subjunctive conditionals with impossible antecedents; second, philosophers already take some of these conditionals to be non-trivially true; and third, such conditionals can be used as counterexamples when the antecedent is not simply the denial of the theory.

2. *What's Right About Impossible Counterexamples*

So far, this paper has considered and rejected various attempts to argue against the use of impossible counterexamples. This subsection attempts to show that there is actually something positively correct about impossible counterexamples, in that there are already reasons to think them probative.

Take the case of ‘utility monsters,’ creatures that derive an enormous amount of utility from objects or states of affairs from which the rest of the world derives less utility. If utilitarianism is true, goes the objection, it is obligatory to render all of one’s holdings to the utility monster. But surely this cannot be the case, so utilitarianism (it is alleged) is false.¹⁰ Notice, however, that crucially, the *possibility* of the utility monster does not appear as a premise in the objection. The most obvious explanation is that the possibility of the monster simply does not matter. After all, no one is arguing (at least, not here) that it is dangerous to *adopt* utilitarianism, since if a utility monster actually appeared, then everyone would have to give up all of their holdings. *That* objection would require the possibility of utility monsters; that is, a cogent reply would have it that utility monsters were impossible. But in contrast, what the objector is doing by offering the utility monster objection is arguing that utilitarianism takes the wrong kind of situation as a decisive reason to render one’s holdings to someone. Suppose somehow that it were discovered (perhaps in the way some

¹⁰ See R. Nozick, *Anarchy, State, and Utopia*, New York, NY: Basic Books, 1974: 41.

philosophers think it is known that zombies are impossible) that utility monsters were impossible. It is unlikely that this would completely save utilitarianism from the utility monster objection; the objector would still insist that the utilitarian was responding to utility monsters the wrong way. Consider a simpler analogue: Suppose a normative ethical theory had it that anyone who saw a square circle was justified in torturing a child in order to derive a modicum of pleasure. It is clear that this theory is false, not simply unevaluable or trivially true or even true. The perception of a square circle provides a counterexample, since even if one were to perceive such an object, one would still be unjustified in torturing a child.

Another example will reinforce this point. In the example of tele-transporters, suppose it were discovered that tele-transporters were actually metaphysically impossible. (Perhaps, as a matter of synthetic necessity akin to water and H₂O, it were learned that nothing that could go through a tele-transporter would be a person with a past. It is unnecessary to go into too much detail.) The point of the ‘malfunctioning tele-transporter’ objection to psychological theories of the persisting self is not that if a tele-transporter *actually* malfunctioned, human beings would *actually* have a difficult time figuring out whom to call identical to the person who entered the tele-transporter. Instead, the problem is that the psychological theory predicts that psychological continuity will always be sufficient for diachronic personal identity, and there is a counterexample to this claim; the psychologist treats the wrong sort of situation as sufficient for persistence of the self.

These examples show that possibility of an antecedent actually plays a minor role—if it plays any role at all—in many important philosophical arguments. Even if the antecedents were discovered to be impossible, philosophers should still treat them as probative, since the possibility of the hypothetical situation simply is not a premise in the argument. Impossible counterexamples may actually be with us already.

III. A FEW IMPLICATIONS

There are several reasons to allow impossible counterexamples in philosophical debate, or at least, to audition them instead of rejecting them out of hand. Suppose philosophers were to allow impossible counterexamples. Would the field of philosophy change significantly?

The previous section already showed that the result would, in some cases, be business as usual. After all, it is not clear how (for example) debates about the persistence of the self over time would change. The field is already troubled by a host of difficult counterexamples, and so the addition of a few more (impossible) ones is unlikely to change matters.¹¹ So there are areas of philosophy that may not be significantly altered with the addition of this new tactic.

However, other subfields are likely to change quite a bit. Return to the example of normative ethics. Generalist normative ethical theories comprise moral principles, generalizations such as ‘one ought not to lie’ or ‘one ought to maximize utility,’ from which one can derive particular judgments. What would be the effect of importing impossible counterexamples? It seems that any generalist theory will suffer from counterexamples, since one can now talk about ‘deontology monsters,’ ‘virtue monsters,’ and a host of other creatures that (say) are morally obligated to lie, act viciously, violate rights or contracts, or whatever. These creatures are probably metaphysically impossible, but again, that is not enough to exclude them from the scope of possible counterexamples. And indeed, these creatures are more like nonexistent Anselmian gods than like logical contradictions, in that no standard language of logic has it as a theorem that deontology monsters do not exist.

Is it open to the generalist to argue that considering such creatures begs the question against the theory? Probably not, since deontology in its simplest form does not state that deontology monsters cannot exist; it simply states that various conditions, such as respecting rights, duties, or contracts, generate certain obligations. What is the point of deontology, after all, if this necessary

¹¹ Compare Parfit 1984, Part 3.

connection between states of affairs and obligations is broken? Another option for the generalist would be to bite the bullet and acknowledge that it really would be obligatory for a deontology or virtue monster (say) to torture an innocent child. But if one allows the biting of bullets this big, it is unclear how to proceed with normative ethical debate in general. After all, the deontologist is likely to criticize the consequentialist for biting similar bullets, as in the Southern sheriff case. If one can just accept such implications willy-nilly, one has almost departed from standard normative ethical debate completely.

These speculations are necessarily brief, given the space allowed here. However, if there are general lessons, one seems to be that some subfields of philosophy would be quite a bit different with impossible counterexamples, and others would remain mostly the same. Another lesson is that perhaps only the (“narrowly”) logically possible (but metaphysically impossible) impossible counterexamples should be considered. Even if so, philosophers should seriously consider impossible counterexamples when formulating or objecting to a theory. The last lesson to draw is that these counterexamples seem to make more of a difference for theories that assert that various conditions produce various *reasons*; compare the case of justified beliefs in square circles, or moral justification in the case of generalism and deontology monsters. Indeed, this observation is strengthened by the realization that once again, the point of ‘utility monster’ style objections is not to criticize theories’ implications about the actual world, but instead to criticize the role of various reasons, or of various conditions as generating various reasons. While descriptive philosophy may be able to downplay impossible counterexamples, normative philosophy such as epistemology and ethics must meet them head-on.