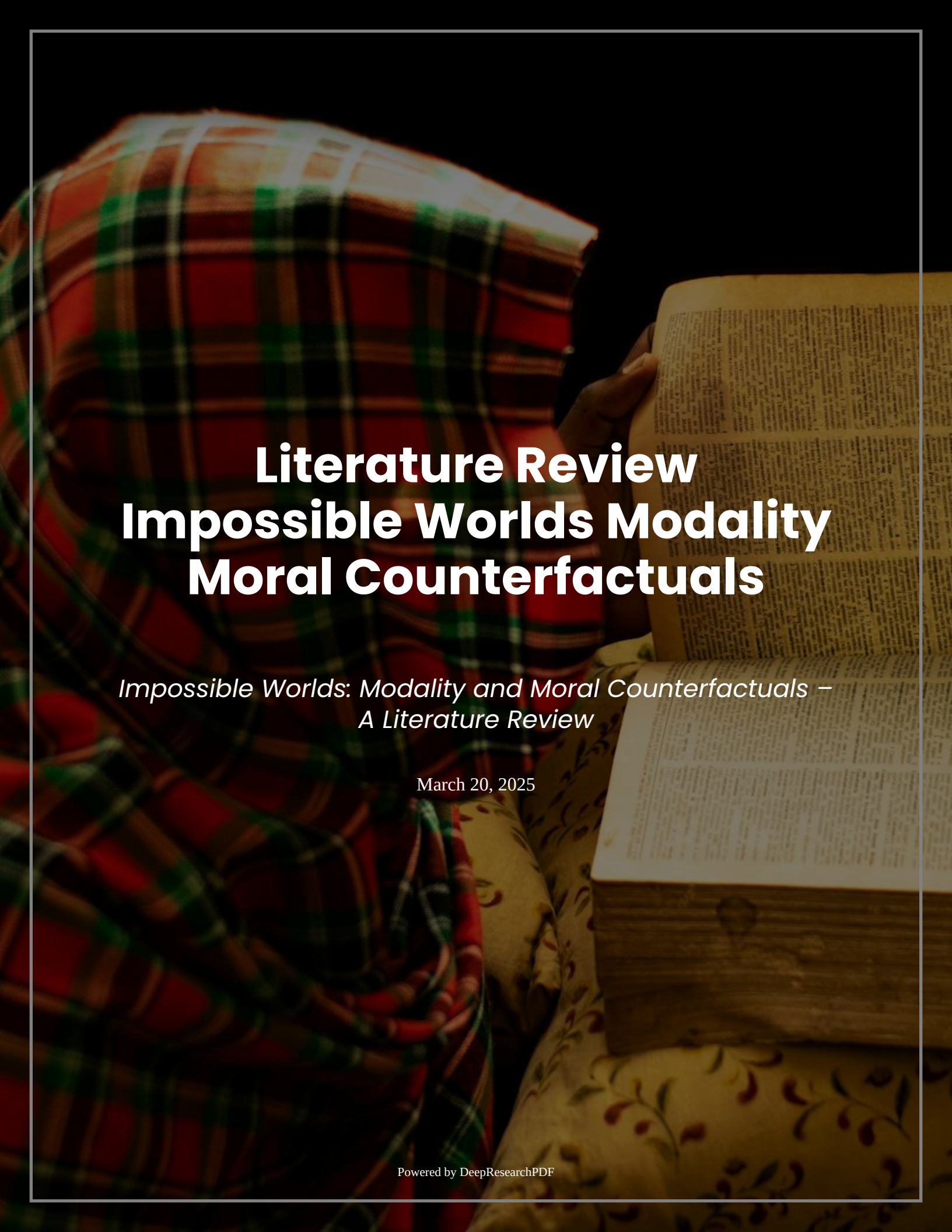

Literature Review Impossible Worlds Modality Moral Counterfactuals

Impossible Worlds: Modality and Moral Counterfactuals – A Literature Review

March 20, 2025

Impossible Worlds: Modality and Moral Counterfactuals – A Literature Review

1. Nature, Existence, and Significance of Impossible Worlds

- *Defining Impossible vs. Possible Worlds:** *Impossible worlds* are commonly described as “ways things *could not* have been,” in contrast to possible worlds as “ways things *could have* been” ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Impossible%20Ways%3A%20just%20as%20possible,Beall%20and%20van%20Fraassen%202003)). In modal metaphysics, a *possible world* is a complete state of affairs that is internally consistent (e.g. Leibniz’s view: a set of compossible things God could have created (Leibniz's Modal Metaphysics - Stanford Encyclopedia of Philosophy)). An *impossible world*, by contrast, is a total description of reality that violates some constraint holding in all possible worlds. For example, if logical laws (like non-contradiction) hold in every possible world, then a scenario where those laws fail qualifies as a logically impossible world ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Logic%20Violators%3A%20another%20definition%20has,a%20world%20in%20which%20the)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Contradiction,not%20in%20the%20fourth%20sense)). Some impossible worlds may violate logical truths (containing outright contradictions ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Contradiction,not%20in%20the%20fourth%20sense))), while others violate non-logical necessities (e.g. a world where a particular metaphysically necessary truth is false, yet without logical contradiction ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=impossible%20world%20in%20the%20first,not%20violate%20any%20logical%20law))). In short, impossible worlds are maximal situations that are not metaphysically or logically possible in the actual sense ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=characterising%20them%20as%20ways%20things,be%20an%20absolute%20impossibility%3B%20and)). This basic conceptual distinction is widely accepted in the literature.

- *Historical Perspectives:** Historically, philosophers often dismissed “worlds” that violate fundamental truths. David Hume famously claimed “the impossible cannot be conceived,” and Moritz Schlick insisted logical impossibilities are literally unthinkable ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=According%20to%20David%20Hume%20%2C,and%20sometimes%20even%20believe%2C%20impossibilities)). G. W. F. Hegel, however, protested that it is a prejudice to assume “the contradictory cannot be imagined or thought,” suggesting we *can* conceive of impossibilities ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=According%20to%20David%20Hume%20%2C,and%20sometimes%20even%20believe%2C%20impossibilities)). The *possible worlds* notion itself traces back to Leibniz: for Leibniz a possible world is any complete concept of creation God could will, excluding contradictions by definition (Leibniz's Modal Metaphysics - Stanford Encyclopedia of Philosophy). Thus, a Leibnizian possible world must be internally consistent; an "impossible world" (a scenario containing a contradiction) would have no reality even in God's mind. For centuries, then, impossible states of affairs were largely treated as *null* or nonsense. Only in the 20th century did philosophers begin seriously considering impossible worlds as a theoretical tool. Early glimmers appear in logic: Stanisław Jaśkowski (1948) and others developed logics for inconsistent systems, implicitly invoking "states" where contradictions hold. N. Rescher and R. Brandom's *The Logic of Inconsistency* (1980) explicitly formulated a "non-standard possible worlds semantics" to handle inconsistency ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=68%3A%20307%E2%80%9327.%20,Grain)), effectively introducing impossible worlds in semantic modeling. The idea gained traction in the 1980s and 90s, with seminal papers like Nathan Salmon's "Impossible Worlds" (1984) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=match%20at%20L87%20these%20ways,Beall%20and%20van%20Fraassen%202003)), Takashi Yagisawa's "Beyond Possible Worlds" (1988), and Daniel Nolan's "Impossible Worlds: A Modest Approach" (1997). These works treated impossible worlds as legitimate objects of inquiry – *ways things could not be* – to address problems that standard possible-worlds theory struggled with.

- *Ontology: Real or Useful Fiction? **A key debate is whether impossible worlds enjoy any genuine ontological status or are merely useful fictions or representations. Advocates of a** parity thesis** argue that "theories of the nature of possible worlds should be

- applied equally to impossible worlds” ([Impossible Worlds | Internet Encyclopedia of Philosophy]). In other words, if one has reason to believe in the existence of possible worlds (either as concrete realities à la David Lewis or as abstract states of affairs), analogous reasons support admitting impossible worlds ([Impossible Worlds | Internet Encyclopedia of Philosophy]). Nolan (2013) observes that the burgeoning uses of impossible worlds mirror those of possible worlds, inviting similar metaphysical treatment ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=This%20entry%20is%20about%20impossibilities,basic%20notions%20of%20worlds%20semantics)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=This%20entry%20is%20about%20worlds,possibility%2C%20however%20this%20is%20further)). On this view, one might countenance impossible worlds as abstract objects (e.g. sets of propositions) or as story-like constructions, extending one’s modal ontology to include “ways things couldn’t be.” For example, an impossible world can be modeled as an inconsistent set of propositions – a maximal but logically inconsistent description of reality. Such an ersatz construction would “exist” as an abstract representation, not a concrete realm ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=,of%20Imagination%E2%80%9D%2C%20Erkenntnis%2C%2082%3A%201277%E2%80%9397)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=,of%20Philosophical%20Logic%2C%2048%3A%20501%E2%80%9321)). This satisfies the parity intuition without positing bizarre concrete impossibilia. Critics, however, balk at full parity. David Lewis, who famously accepted a

plurality of concrete possible worlds, firmly rejected impossible worlds as *literal* worlds. For Lewis, “world” means a possible world; anything impossible is simply not a world at all ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Logic%20Violators%3A%20another%20definition%20has,a%20world%20in%20which%20the)). He and other skeptics hold that impossible worlds are at best *heuristic devices* – bits of semantic bookkeeping – and not genuine elements of reality. One line of argument (a *modus tollens* variant of parity) is: if endorsing possible worlds would force one to endorse impossible worlds (which are unacceptable), then perhaps we should doubt possible worlds too (Impossible Worlds | Internet Encyclopedia of Philosophy). Thus, some philosophers use the absurdity of impossible worlds to *challenge* realist commitments to possible worlds, while others use the utility of possible worlds to *bolster* acceptance of impossible ones (Impossible Worlds | Internet Encyclopedia of Philosophy). The consensus today tends toward a compromise: impossible worlds can be used within semantic or logical theories (thus treated as *ersatz* entities or useful fictions), but their metaphysical status remains contentious. Most who employ them stop short of claiming that “concretely actual” impossible universes exist. Instead, impossible worlds are typically viewed as abstract representations (models, sets of sentences, or functions) that we freely introduce for theoretical purposes ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=,of%20Imagination%E2%80%9D%2C%20Erkenntnis%2C%2082%3A%201277%E2%80%9397)). The question of ontology is still debated in the literature (e.g. Vander Laan 1997’s exploration of the “ontology of impossible worlds” ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=,Without%20Contradiction%E2%80%9D%2C%20Notre%20Dame%20Journal)), and Berto & Jago’s 2019 book defending impossible worlds as legitimate

abstracta ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=,of%20Philosophical%20Logic%2C%2048%3A%20501%E2%80%9321))).

- *Theoretical Significance and Uses:** Why countenance impossible worlds at all? The short answer is their *explanatory utility*. Philosophers found that certain phenomena – especially in logic, language, and philosophy of mind – cannot be adequately modeled using only possible worlds ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=match%20at%20L83%20characterising%20them,be%20an%20absolute%20impossibility%3B%20and)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=,6.6%20Compositionality)). Impossible worlds offer a “more nuanced explanation of modality” and intensional concepts, filling gaps left by orthodox possible-worlds semantics (Impossible Worlds | Internet Encyclopedia of Philosophy) (Impossible Worlds | Internet Encyclopedia of Philosophy). A classic illustration is the problem of *necessarily equivalent propositions*. In possible-world semantics, a proposition is characterized by the set of all possible worlds where it’s true. Any two propositions that are true in exactly the same set of possible worlds are treated as identical in content. This becomes problematic for necessarily false propositions. For example, consider (4) “2+2=5” and (5) “Melbourne both is and is not in Australia.” Neither holds at any possible world (both are false in all possible worlds) (Impossible Worlds | Internet Encyclopedia of Philosophy). Therefore, standard semantics equates (4) and (5) with the *same* possible-worlds intension (the empty set). Intuitively, however, (4) is a false arithmetic claim while (5) is a contradictory geography claim; they differ in subject matter and meaning (Impossible Worlds | Internet Encyclopedia of Philosophy). By introducing impossible worlds, we can distinguish them:

one can stipulate an impossible world where (4) is true but (5) is false, and another where (5) is true but (4) is false (Impossible Worlds | Internet Encyclopedia of Philosophy). This refinement assigns different truth conditions to the two propositions, capturing their distinct content. Thus, impossible worlds serve to avoid conflating distinct impossibilities, yielding a finer-grained (*hyperintensional*) semantics of propositions. This benefit extends broadly: any context where we need to distinguish statements that are necessarily true or necessarily false will potentially require impossible worlds to model subtle differences in content.

Another major motivation comes from the modeling of **intentional states** like belief, knowledge, and imagination. Human beings often entertain or even believe impossibilities (say, a fictional scenario with magic, or an inconsistent worldview). In standard epistemic logic using only possible worlds, an agent who believes a contradiction would trivially believe *everything* (by logical explosion), since any possible-world model that validates an explicit contradiction makes all propositions true. Actual agents, however, can hold inconsistent beliefs without literally believing every proposition. Using impossible worlds, we can model an agent's belief state as a set of *impossible* worlds compatible with everything the agent believes ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=agent%20knows%20,belief)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Impossible%20worlds%20are%20useful%20within,worlds%20models%20usually%20generate%20the)). Because those worlds are not required to obey logical consistency, the presence of a contradiction in the agent's beliefs doesn't make all statements true at those worlds – it only shows the agent's epistemic scenario is impossible, not trivial. In this way, impossible-world semantics can reflect the reality of *logical non-omniscience*: agents do not believe all logical consequences of their beliefs, and sometimes their beliefs are implicitly inconsistent ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=agent%20knows%20,belief)). Work by Hintikka (1975) and later Jago, Bjerring, and others develops impossible-world frameworks to address these issues in epistemic logic, avoiding the unrealistic closure properties of possible-world models ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Impossible%20worlds%20are%20useful%20within,worlds%20models%20usually%20generate%20the)). Impossible worlds also find application in modeling **inconsistent information** (such as contradictory legal codes or scientific theories under development) and in analyzing **fiction and imagination**, where impossible events are commonplace. Graham Priest's short story "Sylvan's Box," for instance, describes a box that is **simultaneously empty and not empty**, an outright impossibility; yet the story is perfectly intelligible (Impossible Worlds | Internet Encyclopedia of Philosophy). To analyze the content of such a fiction, one can invoke an impossible world that makes the story's sentences true, rather than dismissing the story as nonsense. Indeed, theorists of fiction (e.g. in literary semantics) have embraced impossible worlds to represent the content of fantasy and science fiction tales that violate physical or logical laws.

In summary, impossible worlds, though non-actualizable, have become valuable *theoretical entities*. They differ from possible worlds by violating some element of the space of possibilities (logical, metaphysical, mathematical, etc.), and philosophers remain divided on whether they "exist" in any robust sense or are just conceptual aids. Regardless of ontology, their significance lies in the diverse philosophical payoffs: they allow us to reason about counter-logicals and counter-necessities, model cognitive states and discourses that involve impossibilities, and resolve otherwise intractable semantic puzzles. As one general argument puts it, we should accept impossible worlds "because they are useful tools for logicians and philosophers" ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=These%20kinds%20of%20argument%20highlight,Let%E2%80%99s%20look%20at%20some)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=point%20can%20be%20expanded%20into,Let%E2%80%99s%20look%20at%20some)) – their utility in explaining phenomena is the strongest evidence in their favor.

2. Impossible Worlds in Counterpossible Conditionals

- *Counterpossible Conditionals Explained:** A *counterpossible* (or *counterpossible conditional*) is a counterfactual conditional with an antecedent that is impossible – not just false, but necessarily false (in the relevant sense) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=literature%2C%20comes%20from%20counterpossible%20reasoning,trivially%20from%20impossible)). Formally, it has the structure “If A were the case, then B would be the case,” where \(A\) is an impossible proposition. For example, “If Hobbes had squared the circle, then mathematicians would have been amazed” is a counterpossible, since squaring the circle (constructing a square with the same area as a given circle using only Euclidean tools) is mathematically impossible ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=conditionals%20or%2C%20more%20simply%2C%20counterpossibles,These%20include)). Counterpossibles have long been a puzzle for standard theories of counterfactuals. In the familiar Lewis-Stalnaker possible-worlds semantics,

a counterfactual \(A \boxright B\) is true if and only if *all* the closest possible worlds where \(A\) holds are worlds where \(B\) holds ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=match%20at%20L377%20worlds,A)). But if \(A\) is impossible, there are *no possible worlds* where \(A\) is true. The usual semantic treatment in such cases is to deem the counterfactual vacuously true (since the condition on "all \(A\)-worlds" is trivially satisfied when there are no \(A\)-worlds at all) (Williamson on Counterpossibles | Journal of Philosophical Logic). This entails that *every* counterpossible of the form "If [impossible], then X" comes out true, no matter what \(X\) is. In classical logic, indeed, a material conditional with an unsatisfiable antecedent is automatically true. However, many counterpossible conditionals intuitively seem *non*-trivial – some feel true and others false, suggesting we do discriminate between them. For instance, consider two counterpossibles: (6) "If Hobbes had secretly squared the circle, sick children in the mountains of South America at the time would have cared" versus (7) "If Hobbes had secretly squared the circle, sick children in the mountains of South America at the time would not have cared." These have the same impossible antecedent but opposite consequents (Impossible Worlds | Internet Encyclopedia of Philosophy) (Impossible Worlds | Internet Encyclopedia of Philosophy). Our intuitions (as reported by Nolan 1997, who first proposed this example) say that (7) is true – had Hobbes achieved that mathematical miracle in secret, it would have made no difference to sick children in South America, who wouldn't know or care – whereas (6) is false (Impossible Worlds | Internet Encyclopedia of Philosophy). Yet standard semantics would make both (6) and (7) true, rendering the difference unexplained. Likewise, earlier we saw: (1.1) "If Hobbes had squared the circle, then mathematicians would have been amazed" vs. (1.2) "If Hobbes had squared the circle, then mathematicians would not have been amazed." Philosophers generally concur that (1.1) is *true* (mathematicians would indeed be astonished by an impossible feat), and (1.2) is *false* ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=%281,mathematicians%20would%20have%20been%20amazed)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=%281,would%20not%20have%20been%20amazed)). The pair (1.1)/(1.2) exemplifies a **non-trivial counterpossible** – a case where a counterpossible is not vacuously true, since its negation (with the same antecedent and negated consequent) doesn't hold ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Let%E2%80%99s%20call%20a%20conditional%20like,is%20intuitively%20true%2C%20and%20yet)). Such cases undermine the idea that all counterpossibles are uniform or trivial. They show we reason *sensibly* about impossible suppositions: we accept some conclusions and reject others when entertaining an impossible premise ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=literature%2C%20comes%20from%20counterpossible%20reasoning,that%20intuitionistic%20logic%20would%20be)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=not%20only%20false%2C%20but%20impossible,The%20point%20readily%20generalizes)). In other words, we often engage in *counterpossible reasoning*, asking what *would* follow if something impossible were true, and our judgments are not random or all-true – they exhibit consistency with the nature of the supposition.

These considerations motivate alternative semantic approaches. As the Stanford Encyclopedia notes, "the possible worlds semantics for conditionals has trouble accommodating this," even though it works well for conditionals with possible antecedents ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=is%20intuitively%20false,5)). The challenge is to modify or extend the semantics to handle counterpossibles in a non-trivial way, without losing the successes of the Lewis-Stalnaker account for ordinary counterfactuals. Broadly, three strategies emerged in the literature (Impossible Worlds | Internet Encyclopedia of Philosophy):

- **Deny the possible-worlds approach for counterfactuals altogether:** For example, Kit Fine (2012) proposes a semantics for counterfactuals that does not rely on quantifying over possible worlds (Impossible Worlds | Internet Encyclopedia of Philosophy). Fine's framework is *hyperintensional*, distinguishing between different impossible antecedents and assessing counterfactuals via a structured similarity of propositions rather than worlds. This way, one can evaluate counterpossibles directly by logical or conceptual entailment relations, bypassing the need for "nearest A-worlds" when A is impossible (Impossible Worlds | Internet Encyclopedia of Philosophy). Such approaches are complex but offer an alternative to introducing impossible worlds.

- **Accept vacuism (all counterpossibles are true):** This is the orthodox Lewisian view – simply live with the result that counterpossibles have no informative truth-value differences. Timothy Williamson (2007) mounts a contemporary defense of this orthodox stance ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Non,antecedent%20is%20true%20also%20make)). He argues that allowing non-trivial counterpossibles forces one to violate certain desirable logical principles of counterfactuals (discussed more below), so it's better to stick with vacuous truth and treat our intuitions about cases like Hobbes as mislead or pragmatic ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Non,antecedent%20is%20true%20also%20make)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=counterfactuals%20,worlds.%20In%20an)). Proponents of vacuism sometimes explain away apparent counterpossible knowledge as involving tacit assumptions that make the scenario possible, or as using a *suppositional* speech act rather than asserting a genuine conditional (Timothy Williamson, Counterpossibles - PhilPapers). Nonetheless, many find this unsatisfying because it renders debates about impossible hypotheses merely moot or automatically resolved.

- **Extend the Lewis/Stalnaker semantics with impossible worlds:** This third route has been especially popular. The idea is straightforward: when evaluating "if A were the case, B would be the case," allow the set of candidate worlds to include *impossible* worlds in addition to possible ones ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=counterpossibles,This%20motivates%20a%20semantics)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=for%20counterpossibles%20in%20terms%20of,5)). If A is impossible, there are no possible A-worlds, but there may be impossible worlds where A holds. We can then define the truth of the counterfactual in terms of those impossible worlds: $A \boxright B$ is true iff *all the closest* A-worlds (whether possible or impossible) are B-worlds ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=match%20at%20L235%20for%20counterpossibles,5)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=worlds,A)). In effect, this approach changes the domain of discourse from the set of all possible worlds to a broader set that includes impossible worlds, thereby ensuring that even an impossible antecedent can be made true at some “world” and the conditional can be non-vacuously evaluated. Starting with early proposals by Richard Routley (1989) and others in non-classical logic, numerous formal semantic frameworks have incorporated impossible worlds for counterfactuals ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Semantic%20structures%20for%20counterfactual%20conditionals,see%20Vander%20Laan%202004%3B%20we)). Stephen Read (1995), Edwin Mares (1995, 1997), Daniel Nolan (1997), Berit Brogaard & Joe Salerno (2013), Jens Christian Bjerring (2014), and Berto et al. (2018), among others, have all developed or discussed semantics of this kind ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Semantic%20structures%20for%20counterfactual%20conditionals,see%20Vander%20Laan%202004%3B%20we)). Typically, these frameworks are natural extensions of Lewis’s 1973 system: they introduce a class of “impossible worlds” and often a modified similarity relation or selection function to determine which worlds count as the “closest” where \(A\) holds ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Semantic%20structures%20for%20counterfactual%20conditionals,see%20Vander%20Laan%202004%3B%20we)). The truth conditions then mirror the usual counterfactual definition, except that one considers both possible and impossible \(A\)-worlds. This yields the desired results: for Hobbes’s case, an impossible world where Hobbes squares the circle but no distant children care can be considered “closer” (or more appropriate) than any where they do care, making (7) true and (6) false, as intuition demands (Impossible Worlds | Internet Encyclopedia of Philosophy). The approach

successfully differentiates counterpossibles by content of the antecedent and consequent, rather than collapsing them all into trivial truth.

- *Philosophical Motivations and Challenges:** Employing impossible worlds for counterfactuals is motivated by more than just saving our intuitions in toy examples. In many domains of philosophy, we *need* to consider counterfactuals whose antecedents are necessarily false. For instance, in mathematics it is common to ask, “If the axiom of choice were false, what would follow?” Even if one believes the axiom of choice is *mathematically necessary* (true in all possible worlds of set-theory), one can still reason about consequences of its falsehood ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=changes%20the%20example%20to%20%E2%80%9CNothing,%28Field%201989%3A%20237%E2%80%938)). Hartry Field (1989) gives exactly this example, noting that it seems *meaningful* (and non-vacuously true) to say: “If the axiom of choice were false, the Banach–Tarski theorem would fail, cardinal arithmetic would behave differently, etc.” ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=changes%20the%20example%20to%20%E2%80%9CNothing,%28Field%201989%3A%20237%E2%80%938)). This is a counterpossible if one regards Zermelo-Fraenkel set theory + Choice as the only possible mathematics; yet mathematicians and philosophers routinely evaluate such statements productively. Similar reasoning occurs in **metaphysical disputes**. If one metaphysical theory (say, Spinoza’s monism) is regarded by its opponents as necessarily false, critics still often explore **“what if” scenarios** assuming that theory’s truth to draw out its implications ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=,must%20envisage%20situations%20where%20such)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=evaluate%20metaphysical%20theories%20which%20she,we%20have%20made%2C%20impossible%20worlds)). For example, a metaphysician might say: “If Hegel’s Absolute idealism were correct, then history would have a teleological necessity,” even while believing Hegel’s system is fundamentally impossible. These *counterpossible hypotheticals* are used to test or illustrate philosophical positions ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=other%20metaphysical%20debates%20easily%20come,we%20have%20made%2C%20impossible%20worlds)). They would all be trivialized under standard semantics, but with impossible worlds we can treat each metaphysical theory as holding at some impossible world and ask what else is true there ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=other%20metaphysical%20debates%20easily%20come,we%20have%20made%2C%20impossible%20worlds)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=metaphysics%20are%20correct%20and%20wonder,we%20have%20made%2C%20impossible%20worlds)). Another area is **philosophy of science**: one might consider counterlegal conditionals (violating laws of nature or logic) to examine, say, what physics would look like if a certain law were different (some of these are logically impossible if the law is deemed necessary). Impossible worlds give us a formal tool to tackle these questions systematically.

The primary *challenge* in developing an impossible-worlds semantics for counterpossibles lies in defining the comparative similarity (or “closeness”) relation that selects the relevant worlds. Lewis’s framework relied on a clear notion of similarity between possible worlds (guided by weighting of particular facts, etc.). Once impossible worlds enter, how do

we determine which impossible world is “closest” to the actual world, or more generally, how to rank them? The criteria for similarity become less clear when worlds can differ in arbitrary or even inconsistent ways ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Bjerring%202014%2C%20Berto%20et%20al,2)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Lewis%E2%80%99s%201973%20semantics%20for%20counterfactuals,2)). As Vander Laan (2004) discusses at length, fine-tuning the closeness relation to handle impossible antecedents is non-trivial ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Bjerring%202014%2C%20Berto%20et%20al,2)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=counterpossible%20reasoning,see%20Vander%20Laan%202004%3B%20we)). One must balance the **degree of violation** (how severely the impossible world departs from reality’s laws) against preserving relevant facts. For example, to evaluate a counterlogical like “If the law of non-contradiction failed, then X,” one must consider worlds where logic is non-classical. But there might be many such impossible worlds – some that violate just that law and keep everything else as normal as possible, and others that introduce far more chaos. Intuitively, we’d pick the minimal impossibility needed to make the antecedent true. Formal proposals often stipulate that the “closest” impossible worlds are those that are overall most similar to the actual world *except* for the necessary falsehood at hand (they differ only in that respect, and perhaps its direct consequences) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=counterfactuals%20,worlds.%20In%20an)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=trivially%20true,fails%2C%20making%20the%20counterpossible%20false)). Even so, making this rigorous is difficult. Different logicians have suggested various constraints or metrics, but there is not yet a universally accepted method. The **granularity issue** also looms: how fine-grained are impossible worlds? Some systems treat each maximal inconsistent set of sentences as a distinct impossible world, which can lead to a bewildering abundance of worlds (possibly as many as $2^{\aleph_0}$ or higher cardinalities). Deciding when two descriptions yield “the same” impossible world is an ongoing issue ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=,The%20Closeness%20of%20Impossible%20Worlds)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=,6.2%20Defining%20Possibility)). For practical semantic purposes, theorists often individuate impossible worlds just enough to distinguish the cases they care about (e.g. one world for each distinct counterpossible scenario under consideration). Despite these complexities, a significant body of work – including technical developments in relevant logic semantics (Impossible Worlds | Internet Encyclopedia of Philosophy), epistemic logic, and modal semantics – indicates that the impossible-worlds approach can be fleshed out consistently. Nolan’s 1997 framework, for example, demonstrates a non-trivial model for counterfactuals that validates our intuitions about counterpossibles while agreeing with standard semantics on ordinary counterfacturals ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=is%20intuitively%20false,5)). Brogaard and Salerno (2013) further defend such semantics and provide additional examples of counterpossibles that their account handles correctly ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Non,antecedent%20is%20true%20also%20make)).

Before leaving this section, it’s worth noting the *logical cost* of admitting non-vacuous counterpossibles. Williamson (2007) pointed out that certain logical principles must be abandoned if we allow counterpossibles to have substantive truth values ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Non,antecedent%20is%20true%20also%20make)). One important principle is **Strengthening of the Antecedent (or Entailment):** normally, if \(B\) logically follows from \(A\), then “If A were the case, B would be the case” is automatically true. In particular, any *strict implication* (where \(A\) strictly implies \(B\) in all accessible worlds) entails the corresponding counterfactual ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=counterfactuals%20,worlds.%20In%20an)). But suppose \(A\) is impossible and \(A \strictly\rightarrow B\) holds vacuously (since there are no possible \(A\)-worlds where \(B\) fails). A non-vacuous counterpossible semantics might find a closest *impossible* \(A\)-world where \(B\) fails, thus making the counterfactual “If A then B” false ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=counterfactuals%20,worlds.%20In%20an)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=trivially%20true,fails%2C%20making%20the%20counterpossible%20false)). In effect, one can have a case where \(A \strictly\) implies \(B\) (trivially, over possible worlds) but \(A \boxright B\) is false because of an impossible world counterexample. This violates the entailment principle that the strict conditional entails the subjunctive conditional. Another principle that can fail is **Exaggeration** (if \(A \boxright B\), then \((A \land C) \boxright B\)); adding more to an impossible antecedent might change which impossible worlds are considered closest, so a counterfactual could cease being true when the antecedent is strengthened with irrelevant material. These and other logical peculiarities are the trade-offs for a more expressive counterpossible semantics. Williamson and others argue that such deviations are unacceptable and hence every counterpossible should be deemed vacuously true ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Non,antecedent%20is%20true%20also%20make)). Advocates of impossible worlds respond that these “failures” are not catastrophic and in fact align with our intuitions in the contexts where counterpossibles matter (we *don’t* expect many normal logic-laws to seamlessly extend to counterlogicals). The debate remains active, but many contemporary philosophers are willing to jettison some logical principles in exchange for treating counterpossible reasoning as genuinely informative.

In sum, the use of impossible worlds in the semantics of counterfactuals is a robust approach to handling *counterpossible conditionals*. It allows such conditionals to be true or false in a non-trivial way by evaluating them at impossible worlds where the antecedent holds ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Let%E2%80%99s%20call%20a%20conditional%20like,is%20intuitively%20true%2C%20and%20yet)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=is%20intuitively%20false,5)). This approach has been applied to examples in mathematics, metaphysics, theology, and elsewhere, often vindicating common-sense judgments that would otherwise conflict with the vacuist (all-true) result ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=%281,would%20not%20have%20been%20amazed)) (Impossible Worlds | Internet Encyclopedia of Philosophy). The key philosophical motivation is to preserve the meaningfulness of reasoning “under impossible suppositions,” which appears to be an important aspect of how we discuss necessary truths and logical laws ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=literature%2C%20comes%20from%20counterpossible%20reasoning,that%20intuitionistic%20logic%20would%20be)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=to%20reasoning%20about%20entire%20theories,nature%20of%20their%20subject%20matter)). The main challenges are formal (how to systematically implement “closest impossible worlds”) and ideological (whether one is willing to countenance impossible worlds at all). As the next sections will show, these issues take on special significance in the realm of **morality**, where the status of certain truths as necessary or contingent is itself debated, and where counterfactuals play a role in moral reasoning and metaethical arguments.

3. Moral Counterfactuals and Impossible Worlds

- *Moral Necessity and Counterfactuals:** Consider a counterfactual like, “If murder were permissible, then it would be permissible to kill Bob.” On its face, this is a straightforward statement about a hypothetical moral situation: it posits a (currently false) antecedent – that murder is morally permissible – and draws a consequence about a particular case (Bob’s murder being permissible). However, many philosophers (especially in traditions of moral realism or theistic ethics) maintain that fundamental moral truths are not just contingently true but *necessarily* or *immutably* true. For example, one might argue that “Murder is wrong” is true in all possible worlds – it is not merely a fact about our world or conventions, but a truth that could not have been otherwise (akin to a mathematical or logical truth). On such a view, the antecedent “murder were permissible” describes a state of affairs that is **morally impossible**: no possible world exists where murder is morally permissible, assuming the wrongness of murder is an absolute, exceptionless truth. This position is sometimes implicit in robust moral realism and often explicit in theistic ethics (where moral laws reflect the nature of a necessarily good God). For instance, if one believes it is *metaphysically necessary* that cruelty without justification is wrong, then any scenario that makes it right or permissible is an impossible scenario. Some authors clarify the distinction between something being conceptually true vs. metaphysically necessary in this context (Microsoft Word - Necessary Moral Truths and Theistic Metaethics.doc) (Microsoft Word - Necessary Moral Truths and Theistic Metaethics.doc). For the sake of discussion, let’s assume the stance that core moral facts (like “murder is wrong”) are indeed necessarily true. Under that assumption, the counterfactual “If murder were permissible, then it would be permissible to kill Bob” becomes a *counterpossible*: its antecedent is a necessarily false proposition. Standard semantics would then judge it vacuously true (if no possible world permits murder, the conditional holds trivially). But is it *really* just a hollow truth? On analysis, the conditional actually has some intuitive content: it seems to be saying that *in any scenario where the moral rule against murder is absent, that scenario consistently allows individual acts of killing like Bob’s*. In fact, one might argue this particular counterfactual is

- analytically true given the antecedent – it's almost a tautology in deontic terms ("if in some situation murder is permitted by the moral code, then a specific instance of murder would be permitted as well"). So, triviality here might not be problematic; this example likely would be true under any approach. But consider a more nuanced moral counterfactual: **"If it were morally permissible to torture infants for fun, then society would flourish."** This has an impossible antecedent according to a moral realist (torturing infants for fun is necessarily wrong), yet we have the strong intuition that the conditional is *false* – even trying to imagine that perverse moral law, we don't think it would lead to a flourishing society; more likely the opposite. If all counterpossibles are automatically true, we lose the ability to mark such a statement as false or to discuss why it would be false. Thus, moral philosophers face a similar dilemma as in the logical and mathematical cases: moral counterfactuals with impossible antecedents appear in reasoning, and treating them all as vacuous truths can be problematic for analysis.

- *Why consider counterfactuals that contravene moral truths?** There are several reasons such *moral counterpossibles* arise and matter:

- **Philosophical and Metaethical Debates:** In metaethics, one prominent discussion involves whether moral truths are necessary or contingent. Some argue that if moral realism is true, moral facts would be the same in any possible world – for example, any world with pain and pleasure has the moral truth that pain (without outweighing goods) is bad, etc. Critics of this view sometimes use counterfactual tests: "If the moral facts were different, would we still believe them?" to probe the nature of moral knowledge. Interestingly, if moral truths are necessary and counterpossibles are trivially true, then the counterfactual "If moral facts were different, we would have false beliefs" is trivially true – which could *help* the realist answer certain epistemic puzzles by meeting a condition called "sensitivity" (Microsoft Word - Five Kinds of Epistemic Arguments Against Robust Moral Realism for web). (The idea here, outlined by philosophers like David Enoch, is that because the antecedent "moral facts were different" is impossible,

- the conditional is true vacuously, so in a sense our moral beliefs are safe from error across all possible worlds (Microsoft Word - Five Kinds of Epistemic Arguments Against Robust Moral Realism for web). This is a double-edged sword: it defuses one worry about moral knowledge but only by leaning on the vacuity of the counterfactual, which some find unsatisfactory.) More directly, consider debates about **moral objectivity**. Robust objectivists like Enoch note that moral truths seem to behave like objective facts in counterfactuals – e.g. "If our attitudes or culture were different, *would* killing innocents be right?" Most people's intuition (unless they are relativists) is *no* – even if everyone approved of murder, it would still be wrong ([[PDF] Five Kinds of Epistemic Arguments Against Robust Moral Realism*](https://philarchive.org/archive/SCHFKO#:~:text=Realism,that%20if%20the%20moral)). This suggests that the truth of moral claims isn't dependent on our contingencies, reinforcing their necessity. But to say "if everyone approved of murder, it would still be wrong" is itself a counterpossible if one thinks "murder is wrong" is necessary and cannot be overturned by consensus. So again we either declare that statement trivially true or find a way to evaluate it meaningfully. This shows up in discussions contrasting moral truths with mere conventions: objective moral truths are those that* hold their truth* even under contrary suppositions (just as an objective scientific fact, like "smoking causes cancer," remains true in counterfactual scenarios where people doubt it ([[PDF] David Enoch - Why I am an Objectivist about Ethics](https://r.jordan.im/download/philosophy/David%20Enoch%20-%20Why%20I%20am%20an%20Objectivist%20about%20Ethics.pdf#:~:text=%5BPDF%5D%20David%20Enoch%20,than%20like%20purely))).

- **Moral Theory Testing:** Philosophers sometimes test moral theories with hypotheticals that the theory itself says are impossible. A prime example is the **Divine Command Theory (DCT)** in ethics, which states that whatever God commands is morally obligatory. Opponents of DCT often pose the *Euthyphro-style* question: "What if God commanded something horrific (say, torture)? Would it thereby become good/obligatory?" DCT advocates typically respond that God is

- essentially good and *could not* command torture – it's impossible for God to issue a cruel command. The debate then shifts to the counterpossible: *If, per impossible, God did command torture, would it be right?* This is known as the **"counterpossible terrible commands"** scenario ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=,a%20counterfactual%20with%20an)) ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=%3E%20Terrible,act%20would%20be%20morally%20obligatory)). Intuitively, many feel that even if we imagine God commanding torture, it *would not* become morally right – thus, "If God commanded us to torture infants for fun, it would still be wrong to do so" seems plausible ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=Now%2C%20we%20can%20look%20at,the%20DCT%20still%20have%20the)) ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=false%20counterpossible.%20In%20addition%2C%20Sinnott,of%20any%20technical%20details%20about)). However, DCT implies the opposite: under that hypothetical, torture *would* be obligatory (since God's commands define moral duties) ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=counterpossible%20is%20true%3A%20If%20God,paper%2C%20I%20argue%20that%20the)). DCT proponents like Robert Adams and William Lane Craig bite the bullet via *vacuism*: since God's issuing such a command is impossible, the conditional "If God commanded torture, it would be morally obligatory" is vacuously true, and hence not a genuine strike against DCT ([A Non-Vacuist Response to the Counterpossible Terrible

Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=such%20a%20counterpossible%20is%20false,paper%2C%20I%20argue%20that%20the)). They maintain that we aren't conceding anything substantive by allowing that counterpossible, because the antecedent can never obtain (Formulating the Moral Argument | Reasonable Faith) (Formulating the Moral Argument | Reasonable Faith). Critics (e.g. Morriston 2009, W. Sinnott-Armstrong 2009) insist that we *can* coherently consider that scenario and that our moral intuition screams that the conditional is *false* – which, if one agrees, would mean DCT yields a false counterpossible consequence ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=impossible%20for%20a%20necessarily%20morally,a%20counterfactual%20with%20an)) ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=match%20at%20L174%20false%20counterpossible,of%20any%20technical%20details%20about)). This dispute illustrates why moral counterpossibles matter: one side treats them as meaningful tests of a moral theory's implications, while the other side tries to dismiss them as harmless impossibilities. The discussion has prompted sophisticated responses; for instance, Frederick Choo's recent paper (2024) offers a "non-vacuist" defense of DCT, attempting to show that even if we allow the counterpossible to be evaluated non-trivially, the intuition of its falsehood can be addressed without rejecting DCT ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=Critics%20of%20Divine%20Command%20Theory,that%20we%20lack%20reason%20to)) ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=counterpossible%20terrible%20commands%20objection%20fails,a%20new

%20response%20to%20the)). Regardless of who is right, the conversationdemonstrates that moral counterfactuals with impossible antecedents areactively used in moral philosophy to probe the coherence andconsequences of ethical doctrines.

- **Exploring Moral Contingency vs. Necessity:** Counterfactuals are a natural way to ask “Could moral facts be otherwise?” Even outside of specific theories, philosophers wonder whether things like fundamental rights or values could have been different. Are we to say, for example, “If kindness were not a virtue, would it be acceptable to be cruel?” Many hold that basic virtues and vices are necessary – kindness could not *fail* to be a virtue in any world where the concept applies. But thinking through such a counterfactual can help clarify why we consider kindness inherently good. Perhaps we realize that a world where cruelty is celebrated is one with inverted meanings or with agents so unlike us that calling it a difference in *moral truth* might be a category mistake. Indeed, some argue that if a community’s “moral beliefs” are too radically off (e.g. they think wanton cruelty is commendable), we might question whether they are even talking about the same *moral* concepts as us, rather than having an entirely alien normative framework (Microsoft Word - Five Kinds of Epistemic Arguments Against Robust Moral Realism for web) (Microsoft Word - Five Kinds of Epistemic Arguments Against Robust Moral Realism for web). This touches on the **coherence of contrary moral norms**: is an impossible moral world genuinely thinkable, or does it dissolve the very idea of “morality”? Hilary Putnam and others have raised analogous points about conceptual change (e.g. a “world” where water isn’t H_2O might not really be a coherent world, depending on semantics). In moral philosophy, some suggest there are limits to how far moral truths can be altered without changing the subject. But impossible-world talk allows us to simulate even those drastic differences and ask, “what if *morality* as we know it didn’t hold?” This is a double-edged sword: it can generate insight (by isolating what depends on what) but also potentially confuse if we’ve changed the meaning of terms. Philosophers tread carefully, sometimes prefacing such suppositions

- with “*per impossibile*” to indicate a hypothetical truly outside the realm of possibility (as in the DCT debates: “If (per impossibile) God commanded atrocity…” ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=match%20at%20L111%20%3E%20Terrible,act%20would%20be%20morally%20obligatory))).

- *Necessity vs. Contingency of Moral Truths:** Whether moral truths are necessary is itself contested. Naturalist and relativist metaethicists often consider moral facts *contingent*. For example, a moral naturalist might say moral facts supervene on natural facts (facts about human psychology, society, etc.), which could vary by world. In a world with very different natural features or forms of life, what is morally right or wrong might indeed be different – not because the *principles* are different, but because the circumstances to which they apply differ radically (or possibly because the principles themselves could have been different if based on evolved sentiments, etc.). On this view, a statement like “Murder is wrong” might be *generally true* in all worlds where sentient beings feel pain and value life, but it’s not a logical or metaphysical necessity on par with “2+2=4”. Moral truths could be “necessary relative to certain background conditions” but not absolutely necessary. If one holds such a stance, many moral “counterpossibles” cease to be counterpossible at all – they become ordinary counterfactuals about worlds with different conditions. For instance, a utilitarian might ask: “If human nature were such that people felt no empathy, would killing innocents still be wrong?” If empathy is a key factor in why killing is wrong, a world of creatures without empathy might be a possible world where moral truths are altered. There are science-fiction explorations of this: e.g., if an alien species had completely different sensibilities, would our moral truths apply to them or would they have a different morality? These are *possible* worlds from a broad perspective, although from our viewpoint they might seem “impossible” if we assume our morality is universally binding. The **analytic value** of such moral counterfactuals lies in

- testing the robustness of moral principles: are they invariant or do they depend on certain parameters? If one finds that even imagining extreme changes, one cannot shake the truth of a moral principle, that suggests a kind of necessity or universality. If instead one can imagine a coherent world where that moral principle doesn't hold (and things maybe function differently yet not chaotically), that suggests the principle might be contingent. For example, some philosophers argue that certain basic moral norms (like prohibitions on harm) might actually be *contingent on human nature* – had we been very different creatures, what counts as "harm" or whether harm is bad might differ. Others, like Kantian ethicists, might say the structure of rational agency ensures some moral laws (like respect for persons) in any possible world with rational agents. This debate is often implicitly carried out with counterfactual reasoning, even if the term "impossible" isn't always used.

- *Are Moral Counterpossibles Meaningful?** There is a worry about coherence: if one truly believes a moral truth is necessary, when we utter a counterfactual supposing its negation, are we talking about the *same* moral concepts or sneaking in a change of meaning? Some contend that counterpossibles in morality may lose grip on the concept. For instance, saying "If cruelty were morally good, then ..." might be confusing because part of what "cruelty" means to us is bound up with its badness. If an interlocutor protests that the scenario is inconceivable ("I can't imagine a world where gratuitous cruelty *is* good"), they echo Hume and Schlick's sentiment about inconceivability of impossibilities, but now in the moral domain. Yet, as we saw, Hegel would encourage pushing beyond that prejudice and at least trying to imagine the "contradictory" case ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=According%20to%20David%20Hume%20%2C,and%20sometimes%20even%20believe%2C%20impossibilities)). Philosophers do attempt it, often for heuristic value: exploring impossible moral worlds can clarify why our actual moral truths hold. It can serve as a reductio: assume the opposite of a cherished

moral principle and see what follows; if the results are abhorrent or nonsensical, that reinforces the necessity or strong justification of the original principle. This is analogous to reductio ad absurdum in logic, which some argue implicitly involves impossible worlds to make sense of deriving a contradiction from an assumption (Impossible Worlds | Internet Encyclopedia of Philosophy).

In practice, moral counterpossibles are frequently used as *thought experiments or argumentative tools*. The earlier example of Divine Command Theory is a case in point: the impossible scenario of God commanding evil is used to test the theory's mettle. Another example: one might argue for the immutability of moral principles by saying *"Even if might made right (i.e. if whoever is strongest determined morality), it would still be wrong to kill the innocent just for power."* This statement is counterpossible if we assume "might makes right" is not actually possibly true (because moral truth isn't determined by power in any possible world). Articulating it, however, communicates the speaker's conviction that moral rightness doesn't depend on power – a way of affirming moral objectivity. The statement's truth on the face of it would be vacuous (since the antecedent is impossible given their stance), but its meaning and use are not pointless; it emphasizes a key feature of our moral concept: that it isn't reducible to brute force.

- *Moral Necessity and Analyticity:** Some moral truths might be considered *analytic* or true by definition (e.g. "One ought to pursue the good" might be seen as part of what "ought" means). Thomas Aquinas's maxim "do good and avoid evil" is often cited as a logically or conceptually true principle (Microsoft Word - Necessary Moral Truths and Theistic Metaethics.doc). If something is an analytic moral truth, then a counterfactual negating it could even be seen as *meaningless* or self-contradictory, not just impossible. However, most substantive moral claims (like "murder is wrong") are not mere tautologies; they convey content about the world (murder, harm, persons, etc.). The necessity, if any, is synthetic – arising from the nature of persons, or God's nature, or rationality, etc. Thus, "murder is wrong" could be necessary yet still *informative*, and "if murder were permissible" is not a contradiction in

- terms, but rather an attempt to contemplate a world governed by an inverted moral law. The analytic truth case is rare; more often philosophers treat basic moral truths as synthetic a priori truths (in Kantian terms), which are necessary in a broad sense but not trivially so. Therefore, they allow that we can talk about their negations, albeit as impossibilities, to see what that entails.

To summarize this section: Moral counterfactuals like “If X were (morally) permissible, then Y” become *counterpossible* when one presupposes that X is necessarily impermissible. Such statements are useful in moral reasoning (to test theories, illustrate objectivity, etc.), but they confront us with the same logical-semantic issues discussed earlier. Our intuitions often provide non-vacuous truth values to these conditionals (some we think would be true, others false), which puts pressure on semantic theories to accommodate that or explain it away. The assumption that moral truths are necessary is central here – if one drops that assumption and allows moral variability across worlds, these conditionals can be treated with standard possible-worlds semantics (though even then, if one holds a very rigid form of supervenience or moral law, some antecedents might still be impossible in a restricted sense). The tension between the *apparent meaningfulness* of moral counterpossibles and the *strict necessity* of moral truths is a focal point of recent scholarship, leading to efforts to reconcile the two.

4. Reconciling Impossible-Worlds Semantics with Moral Necessity

Given the above, how do philosophers reconcile the use of impossible worlds (or non-vacuous counterpossibles) with the idea that moral truths are necessary? There are a few general approaches:

- *(a) Embrace Vacuism – Trivial Truth of Moral Counterpossibles:** One straightforward strategy is to maintain that yes, fundamental moral truths are necessary, and therefore any counterfactual assuming their falsity is vacuously true (or at least not false). This was essentially

- Robert Adams's move regarding the "God commands cruelty" scenario: since God's goodness is necessary, "If God commanded cruelty, it would be right" is not actually a damaging implication but a vacuous truth inside an impossible supposition. Williamson's general defense of vacuism would extend here: our feeling that *"if, per impossibile, cruelty were good, society would collapse"* is just the expression of a heuristic, not a literally evaluable conditional (Timothy Williamson, Counterpossibles - PhilPapers). The vacuist stance can preserve the sanctity of moral necessity by refusing to compare actual morality with "alien" moral frameworks – there is no actual comparison to be made, so any statement bridging them lacks substantive truth conditions. However, vacuism has the drawback of undercutting the informative value of moral counterfactuals. It would imply that debates like the DCT objection are misguided (the counterpossible premise "If God commanded X…" has no real truth value beyond triviality). While some are content with that (arguing the objection indeed "concedes too much" by granting the impossible supposition (Formulating the Moral Argument | Reasonable Faith) (Formulating the Moral Argument | Reasonable Faith)), others see it as dodging the challenge. Still, a number of philosophers stick to vacuism for counterpossibles across the board – in effect, they accept that necessary moral truths make certain "what if" questions moot. This stance keeps the logic of counterfactuals simple (no need for impossible worlds) and avoids any hint of relativizing morality to circumstances.

- *(b) Non-Vacuist Semantics via Impossible Worlds:** Another approach is to accept that we *do* sometimes need to evaluate moral counterpossibles non-trivially and thus apply the impossible-worlds semantics to the moral domain. This means allowing impossible worlds where, say, murder *is* permissible, and examining what holds in those worlds. The key is to do this in a controlled way that doesn't undermine the idea that murder is wrong in all *possible* worlds. In technical terms, we distinguish between *metaphysical* possibility and a broader logical or conceptual possibility: a world where murder is permissible is not metaphysically possible (given moral necessity), but we treat it as a

- logically conceivable world (no outright contradiction in the description “a world with different moral laws”). It is an impossible world of the first kind (violating a non-logical necessity) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=impossible%20world%20in%20the%20first,not%20violate%20any%20logical%20law)). We then evaluate the counterfactual with respect to such worlds. For example, to evaluate “If murder were permissible, Bob’s murder would be permissible,” we consider the closest impossible world(s) where murder is permissible and check the status of Bob’s murder. Almost by definition, in those worlds Bob’s murder is indeed permitted (since that world presumably has the general law that allows murder), so the conditional comes out true – matching our intuition. For a more substantive case, “If cruelty for fun were morally acceptable, society would flourish,” we consider an impossible world where cruelty-for-fun is morally acceptable. We also need to fill in other details of that world in a plausible way – perhaps holding as much of the actual world constant as possible except that one moral rule is inverted. In such a world, would society flourish? Likely not: perhaps trust and cooperation break down, etc. If in all reasonably close such worlds society does *not* flourish, then the counterfactual is false, as we expect. This way, impossible-world semantics yields sensible truth conditions: it allows the antecedent to *engage* with consequences in a way impossible under vacuism. The challenge, as always, is defining “closeness” or the selection of the right impossible worlds. We typically would allow only the *minimal* change to morality needed to make the antecedent true, and assume other natural facts remain as they are. In the cruelty example, we’d take an impossible world where the only difference is that the moral code is inverted on cruelty, but human nature, psychology, etc., remain the same as actuality. In that world, cruelty being permissible likely leads to lots of pain (a natural fact), which by our own moral lights is bad, but in that world’s moral system is fine – yet the question of society flourishing is factual, not moral, so we can evaluate it by our standards of societal health. If one is consistent, one might say even “flourishing” might be defined differently

in that world, but usually we project our meanings into the scenario to draw conclusions. Philosophers like Nolan, and more recently Kocurek (2021) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=match%20at%20L377%20worlds,A)), have argued that adopting impossible worlds for counterpossibles (including moral ones) is overall more plausible than maintaining vacuism, because it preserves the meaningful distinctions speakers make. The *cost* is that we accept moral truths as necessary but still countenance *impossible worlds* that violate them, a seemingly strange move: it's like saying "Murder is wrong in every possible world, but we can still quantify over 'worlds' where it's not wrong." This is acceptable if one's view of worlds is semantic (they are representational devices, not realms of being). A moral realist can thus maintain: in reality, murder could never be right, but as a *conceivable supposition*, we can represent that situation and analyze it. This reconciliation hinges on a kind of fiction: treating the impossible scenario "as if" it were a world to learn something about our actual morals. As long as one is clear that these are *not* genuine alternative realities, it doesn't commit one to anti-realism. It's analogous to mathematicians exploring non-Euclidean geometries – logically consistent systems that are "impossible" in Euclidean axiom terms but by studying them you learn structure and perhaps even something about Euclidean geometry by contrast.

- *(c) Revisiting Moral Necessity:** A different tactic is to question whether moral truths are in fact *strictly* necessary. Some philosophers are open to the idea that at least some moral truths are contingent. If, for instance, one adopts a form of moral naturalism where moral properties are anchored in natural properties, then a sufficiently different natural world could yield different moral truths. In that case, many moral "counterpossibles" aren't impossible at all – they refer to possible worlds with different conditions. For example, one could argue "If humans had no capacity for empathy, perhaps murder wouldn't be deemed wrong (because the basis for its wrongness—empathy for victims—would be absent)." That becomes an ordinary counterfactual

- about a possible world where human nature is radically altered. Of course, this approach can be controversial, as it might conflict with strong moral realist intuitions. But it's a way to reconcile analyzing such counterfactuals without needing impossible worlds: deny the premise that makes them counterpossible. In divine command ethics, a parallel would be to deny that God's nature *absolutely* forbids the scenario – maybe say God, being omnipotent, could command anything (though that defies the usual view of God's essential goodness). Very few theists would go that route; instead, they prefer vacuism (as above) or the impossible-worlds analysis. In secular ethics, some philosophers (especially those influenced by evolutionary biology or anthropology) see morality as an adaptive or cultural system that *could* have been otherwise to some extent. They might consider "counterpossible" moral scenarios as simply exploring the space of possible cultures or psychologies, not violating any true necessity. Thus, the tension is resolved by demoting moral necessities to very strong contingent truths. However, this move has its own costs: it may undermine claims of objective moral truth being exceptionless and eternal. The literature often circles back to whether we have reason to view moral truths like logical truths or not – a deep question beyond pure semantics.

- *(d) Hybrid Approaches and Pragmatics:** Some have suggested a pragmatic or metalinguistic solution: when we utter a counterpossible like "If murder were permissible, then X," perhaps we are implicitly shifting perspective or speaking within a *pretend* context. On this account, we don't literally assert a proposition about an impossible world; rather, we engage in a kind of *make-believe* reasoning or a conditional intention. In the case of the DCT example, one could say the atheist is speaking *within the fiction* of a godless universe or a capricious God to see what morality would entail. Within that fictional context, ordinary possible-world reasoning applies (since within the fiction, maybe it's possible that God commands evil). When the theist hears it, they translate it differently. This perspective can be couched in terms of *story operator* ("Suppose for the sake of argument that murder is permissible; then…") as Sendłak (2021) discusses (Counterpossibles,

- consequence and context - Taylor and Francis). The advantage is that one might preserve classical logic by saying these aren't genuine counterfactuals in the semantic sense, but rather *suppositional discourses* governed by conversational rules. The disadvantage is that it's less of a systematic semantic theory and more of a case-by-case pragmatic interpretive tool. Still, it highlights that often when we use counterpossibles, especially in moral and philosophical debates, we are inviting the listener into a kind of *what-if game* rather than stating a fact about another world.

- *Key Figures and Current Debates:** Many of the key thinkers have been mentioned: Daniel Nolan's work (1997, 2013) is foundational for advocating impossible worlds and surveying the landscape ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=A%20look%20at%20the%20rapidly,reduced%20to%20four%20main%20items)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=%281,mathematicians%20would%20have%20been%20amazed)). Graham Priest has consistently argued for embracing "logic violator" impossible worlds, especially to handle paradoxes and even to question the absoluteness of logical laws – though Priest's focus is often more on metaphysical and logical impossibilities than moral ones ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Logic%20Violators%3A%20another%20definition%20has,a%20world%20in%20which%20the)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Classical%20Logic%20Violators%3A%20another%20definition,impossible%20in%20the%20third%20sense)). Timothy Williamson stands as a prominent defender of the orthodox view (all counterpossibles are trivially true) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Non,antecedent%20is%20true%20also%20make)); his arguments in *The Philosophy of Philosophy* (2007) and more recently *Suppose and Tell* (2020) lay down challenges that any non-vacuist theory must meet. Berto, Jago, and their collaborators (2018, 2019) explicitly address Williamson's objections, refining impossible-worlds semantics to avoid triviality without inconsistency ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=,of%20Philosophical%20Logic%2C%2048%3A%20501%E2%80%9321)). Their 2018 paper "Williamson on Counterpossibles" responds to his criticisms point by point ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=,of%20Philosophical%20Logic%2C%2048%3A%20501%E2%80%9321)). In the realm of metaethics, discussions of "necessary moral truths" appear in the context of moral realism and theistic ethics. Figures like C. Stephen Evans, Mark Murphy, or William Lane Craig (theistic) will typically assert moral necessities and default to vacuism or special pleading in counterpossibles (Formulating the Moral Argument | Reasonable Faith). Non-theistic realists like Russ Shafer-Landau or Derek Parfit also believe in robust moral truths, but they seldom address counterpossibles explicitly in their works – it's more an implicit stance. However, some metaethicists, like Peter van Inwagen or Philip Quinn, have touched on "what if God commanded differently" from a theistic perspective, usually siding with "that's impossible, so the hypothetical tells us nothing (or tells us only about our concepts)". In contrast, utilitarians or relativists might be comfortable saying moral truths are contingent, thus eliminating the need for impossible moral worlds.

Currently, there is vibrant interdisciplinary interest in counterpossibles: for instance, experimental philosophers have even begun studying how

people evaluate counterfactuals with impossible antecedents (including moral ones) to see if non-vacuity is a cognitive reality. Early results suggest laypeople *do* distinguish among counterpossibles, treating some as true and others as false, much as philosophers do with Hobbes's circle-squaring example (Impossible Worlds | Internet Encyclopedia of Philosophy). This empirical angle reinforces the philosophical trend toward non-vacuism.

- *Unresolved Issues:** Several issues remain open for further inquiry:

- *Metaphysical Status:* The question "What are impossible worlds?" is still debated. Are they linguistic constructions, abstract propositions, fictions, or something else? The answer might impact how comfortable one is using them in serious moral metaphysics. If one is a staunch realist who wants everything in one's ontology to be "real," one might resist using impossible worlds unless they can be given a respectable ontological status (perhaps as maximal inconsistent *but* structured propositions, as Kment (2014) attempts ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Counterpossible%20reasoning%20may%20also%20show,collections%20of%20structured%20Russellian%20propositions))). If they are just convenient fictions, some worry this might infect the analysis with anti-realism. So squaring the use of a "fictional" device with a realist stance on moral truth is an ongoing subtlety.

- *Closeness Criteria:* Particularly for moral counterpossibles, what makes one impossible world "closer" than another? Is it that it deviates less from the actual moral truths? Perhaps a world where only one moral principle is flipped is closer than one where ten principles are flipped. And how do we measure differences in moral codes? This could lead into exploring a *metric on moral systems* – a new avenue connecting ethics with modal logic. Vander Laan (2004) raised the general issue for impossible worlds; applying it specifically to moral worlds might require thinking about the structure of moral theories and their inter-relations.

- *Moral Epistemology:* If moral truths are necessary, how do we know them? Some argue that if they were necessary, they'd be *a priori* or self-evident. Others say we know them empirically or via intuition. The role of counterfactuals in moral epistemology is interesting: one might test one's grasp of a moral claim by seeing what one would say under contrary-to-fact scenarios (kind of a "modal security" test). If our beliefs would remain firm under any counterfactual perturbation, does that signal anything about their justification? Work by Justin Clarke-Doane and others on "modal security" for moral beliefs ties into this: the idea that for a belief to be knowledge, it should hold not just in the actual world, but in nearby possible worlds. But if moral truths have no *nearby* worlds where they fail (only remote impossible ones), how do we apply such criteria? This is still being figured out (Microsoft Word - Five Kinds of Epistemic Arguments Against Robust Moral Realism for web) (Microsoft Word - Five Kinds of Epistemic Arguments Against Robust Moral Realism for web).

- *Analytic vs. Substantive Counterpossibles:* Not all counterpossibles are equal. Some are "conceptual truths" that if violated might yield analytically false consequents. For example, "If good were evil, then nothing would matter" – here the antecedent is nearly incoherent. Others are substantive: "If pain were pleasant, then life would be very different" – here we can chisel out a scenario more concretely. Sorting which moral counterpossibles give meaningful information and which are too conceptually confused is an area for further philosophical disentanglement.

- *Relevance to Normative Theory:* Do impossible worlds have a role in normative ethical theory or only in metaethics? One could imagine using impossible worlds in, say, deontic logic to handle conflicting obligations (an "impossible world" where you honor both conflicting duties might represent an inconsistent code used to analyze dilemmas). In fact, deontic logicians sometimes use techniques akin to impossible worlds to handle deontic paradoxes (like situations of moral dilemmas or inconsistent norms). This overlaps with paraconsistent logic and is a potential avenue where the two topics of this review meet directly:

- modeling moral dilemmas might require an impossible world (if two moral rules truly conflict, there is no possible world where both are satisfied, but one might use an impossible world to represent the scenario of both being *demanded*).

In conclusion, impossible worlds have become an influential concept in contemporary analytic philosophy, offering solutions to problems in modal semantics and beyond. They differ from possible worlds by encompassing "ways things couldn't be," and while their ontological standing is debated, their utility is evidenced in a range of applications – notably in handling counterpossible conditionals. Moral counterfactuals illustrate a special case where the content under discussion (moral law) might itself be necessarily fixed, creating a prima facie tension with the idea of considering it otherwise. Philosophers have navigated this by either denying the meaningfulness of the contrary hypothesis or by carefully using impossible worlds to explore it without undermining the necessity in the actual sense. Key figures like Lewis, Priest, Nolan, Williamson, Berto, Jago, and many others have contributed to framing the issues, some arguing against and many for the inclusion of impossible worlds in our theoretical toolkit. The debate is far from settled: as our understanding of modality, logic, and morality deepens, the role of impossible worlds is continually refined. What is clear is that they provide a powerful lens – one that allows us to examine not just how things *could* be different, but even how things *couldn't* be, and to glean insight from that exercise.

Bibliography

- **Stanford Encyclopedia of Philosophy – "Impossible Worlds" (2013/2023).** A comprehensive survey by Francesco Berto (revised 2023) covering definitions of impossible worlds, motivations for introducing them, applications (intentional states, inconsistent information, fiction, propositional content, counterpossible reasoning), metaphysical questions, and objections ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Impossible%20Ways%3A%20just%20as%20possible,Beall%20and%20van%20Fraassen%202003)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=conditionals%20or%2C%20more%20simply%2C%20counterpossibles,These%20include)). Provides extensive references to the literature and situates contemporary debates (e.g. the utility vs. logical cost of impossible worlds) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Let%E2%80%99s%20call%20a%20conditional%20like,is%20intuitively%20true%2C%20and%20yet)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Non,antecedent%20is%20true%20also%20make)).

- **Internet Encyclopedia of Philosophy – “Impossible Worlds” (2019).** An accessible overview discussing the *parity thesis* (the idea that arguments for possible worlds equally support impossible worlds) (Impossible Worlds | Internet Encyclopedia of Philosophy) and the theoretical applications of impossible worlds (Impossible Worlds | Internet Encyclopedia of Philosophy). Includes illustrative examples such as distinguishing \(2+2=5\) from a geographical contradiction via impossible worlds (Impossible Worlds | Internet Encyclopedia of Philosophy), and notes on reductio ad absurdum reasoning (Impossible Worlds | Internet Encyclopedia of Philosophy). Summarizes key uses and acknowledges limitations of impossible worlds.

- **Leibniz’s Modal Metaphysics (SEP).** Background on Leibniz’s conception of possible worlds as combinations of compossible individuals (Leibniz's Modal Metaphysics - Stanford Encyclopedia of Philosophy). Highlights the historical notion that “worlds” must be

- internally consistent (ruling out impossible ones in Leibniz's framework). Useful for contrasting modern notions of impossible worlds with classical views on modality.

- **Rescher, N. & Brandom, R. (1980). *The Logic of Inconsistency*.** One of the first full treatments of reasoning with inconsistency using a "non-standard possible worlds" semantics ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=68%3A%20307%E2%80%9327.%20,Grain)). They introduce worlds that are not required to obey consistency, effectively an early theory of impossible worlds in logical space. This work set the stage for later formal developments in paraconsistent logic and relevant logic.

- **Salmon, N. (1984). "Impossible Worlds."** A short article in *Analysis* (44:114–17) explicitly introducing the term "impossible worlds" in the analytic tradition ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=match%20at%20L87%20these%20ways,Beall%20and%20van%20Fraassen%202003)). Salmon argues for treating some statements as true at impossible "ways things couldn't be," foreshadowing later hyperintensional semantics.

- **Yagisawa, T. (1988). "Beyond Possible Worlds."** *Philosophical Studies* 53: 175–204. Explores the idea of extending modal realism beyond the merely possible. Yagisawa suggests ways to meaningfully talk about impossible circumstances (coining terms like "genuine impossible worlds"). Often cited for early advocacy of impossible entities alongside possible ones.

- **Nolan, D. (1997). "Impossible Worlds: A Modest Approach."** *Notre Dame Journal of Formal Logic* 38(4): 535–572. A highly influential paper offering a formal semantics for counterfactuals that incorporates impossible worlds ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=conditionals%20or%2C%20more%20simply%2C%20counterpossibles,These%20include)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=%281,would%20not%20have%20been%20amazed)). Nolan provides the famous Hobbes circle-squaring examples and argues that allowing impossible worlds yields sensible truth values for counterpossibles ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Let%E2%80%99s%20call%20a%20conditional%20like,is%20intuitively%20true%2C%20and%20yet)). He addresses objections and calls the approach “modest” because it doesn’t require committing to concrete impossible realities, only to a broadened semantic space.

- **Nolan, D. (2013). “Impossible Worlds.”** *Philosophy Compass* 8(4): 360–372. A useful survey article summarizing the state of play as of 2013 ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=This%20entry%20is%20about%20impossibilities,basic%20notions%20of%20worlds%20semantics)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=This%20entry%20is%20about%20worlds,possibility%2C%20however%20this%20is%20further)). Nolan reviews various definitions of impossible worlds, their uses in logic and language, arguments for and against their existence, and unresolved issues. Good introductory reading that cites key sources up to that date.

- **Vander Laan, D. (1997). “The Ontology of Impossible Worlds.”** *Notre Dame Journal of Formal Logic* 38(4): 597–620. Examines what

- sorts of entities impossible worlds could be, if we admit them ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=,Without%20Contradiction%E2%80%9D%2C%20Notre%20Dame%20Journal)). Vander Laan presents the “argument from ways” – if worlds are ways things could be, then consider ways things could *not* be ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=1,Worlds)) – and explores how to integrate impossible worlds into one’s ontology (leaning towards an abstract, set-of-propositions view). This is a foundational piece for the metaphysics of impossible worlds.

- **Vander Laan, D. (2004). “Counterpossibles and Similarity.”** In *Lewisian Themes*, eds. Jackson & Priest (OUP). An in-depth discussion of the similarity metric problem when impossible worlds are involved ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Bjerring%202014%2C%20Berto%20et%20al,2)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=counterpossible%20reasoning,see%20Vander%20Laan%202004%3B%20we)). Vander Laan analyzes how to meaningfully rank worlds that violate different laws or truths. An important resource for anyone developing an impossible-worlds semantics for counterfactuals.

- **Priest, G. (1997b). “Sylvan’s Box: A Short Story and Ten Morals.”** *NDJFL* 38(4): 573–581. A whimsical yet philosophical fiction about an impossible object (a box that is empty and not empty) (Impossible Worlds | Internet Encyclopedia of Philosophy). Priest uses it to illustrate points about paraconsistent logic and impossible situations. The “ten morals” provide insight into the philosophical lessons one can draw from considering an impossible scenario. Demonstrates the

- intelligibility of contradictions in a narrative, supporting the view that we can conceive impossibilities ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=According%20to%20David%20Hume%20%2C,and%20sometimes%20even%20believe%2C%20impossibilities)).

- **Priest, G. (2005). *Towards Non-Being: The Logic and Metaphysics of Intentionality*.** (OUP; 2nd ed. 2016). Priest's book-length treatment of noneism (the view that we can refer to non-existent objects) and intentionality, which involves reasoning about impossible objects and situations. Develops a formal semantics with possible and impossible worlds to handle intentional states (belief, etc.) that can be about impossibilities. While not focused on morality, it's a key work for the logical underpinnings of impossible worlds.

- **Brogaard, B. & Salerno, J. (2013). "Remarks on Counterpossibles."** *Synthese* 190: 639–660. Contemporary defense of non-vacuous counterpossibles ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Non,antecedent%20is%20true%20also%20make)). They present linguistic and logical arguments that some counterfactuals with impossible antecedents are false (thus not all are true), challenging Williamson's stance. They also discuss how certain logical principles (like contraction or simplification of antecedents) fail in a non-vacuist setting, but argue this is acceptable. Offers numerous examples and engages with earlier literature (including their short 2007 piece "Why Counterpossibles are Not Trivial").

- **Williamson, T. (2007). *The Philosophy of Philosophy*.** (Blackwell). Chapter 5 of this book is often cited regarding counterfactuals with necessary falsehoods as antecedents ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Non,antecedent%20is%20true%20also%20make)). Williamson articulates the orthodox position that counterpossibles are trivially true, using the collapse of certain inference principles as evidence against non-vacuist approaches ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=Non,antecedent%20is%20true%20also%20make)) ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=counterfactuals%20,worlds.%20In%20an)). He suggests that our intuitions about examples can be explained by conversational implicature or by the mistaken assumption that the antecedent is only *metaphysically* impossible but perhaps logically possible, etc. Williamson's challenge spurred many responses.

- **Williamson, T. (2020). *Suppose and Tell: The Semantics and Heuristics of Conditionals*.** (OUP). Although primarily about conditionals in general, Williamson doubles down on his view about counterpossibles here, and also distinguishes the heuristic use of counterfactual reasoning in mathematics and philosophy from semantic truth-conditions. He acknowledges that people reason about "impossible" things, but maintains that in terms of truth conditions, those counterfactuals are vacuous – it's just that we can glean heuristic info from the process of supposing an impossibility. Provides a sophisticated take on why the semantics can be simple (all true) even if the practice is not trivial.

- **Berto, F., French, R., Priest, G., & Ripley, D. (2018). "Williamson on Counterpossibles."** *Journal of Philosophical Logic* 47(4): 693–713. A direct response to Williamson's arguments ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=,of%20Philosophical%20Logic%2C%2048%3A%20501%E2%80%9321)). The authors, all proponents of non-classical logics or impossible worlds, analyze the logical principles Williamson wants to preserve and show how one can develop a semantics where counterpossibles are non-trivial yet we don't get explosion or incoherence. They address the entailment from strict implication to counterfactual, etc., providing a model that invalidates those principles in a controlled manner. A crucial paper for understanding the current state of the debate.

- **Berto, F. & Jago, M. (2019). *Impossible Worlds*.** (Oxford University Press). A book-length exploration of the metaphysics and logic of impossible worlds by two leading experts. Covers the motivations, different conceptions (they discuss granularity and composition of worlds), and applies impossible worlds to various philosophical puzzles. Likely includes discussion on counterpossible conditionals, intentionality, and perhaps touches on normative domains. This book is a comprehensive resource consolidating much of the recent work on the topic.

- **Kocurek, A. (2021). "Counterpossibles."** *Philosophy Compass* 16(10): e12787. A recent survey focusing specifically on counterpossible conditionals ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=match%20at%20L377%20worlds,A)). Kocurek reviews linguistic data, theoretical positions (vacuism vs non-vacuism), and various semantic frameworks. A valuable up-to-date summary that also connects to related issues like epistemic counterfactuals and the role of context. Good for seeing the landscape of arguments and where consensus (if any) lies.

- **Morriston, W. (2009). "What if God Commanded Terrible Things?"** (in *Philosophia Christi* 11). Although not directly cited above, this paper is the source of the DCT "terrible commands" objection in modern form. Morriston argues that DCT implies an unacceptable counterpossible: *If God commanded something terrible, it would be*

- *obligatory*. He contends this is false, thus rejecting DCT. Choo (2024) cites Morriston's formulation ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=Now%2C%20we%20can%20look%20at,the%20DCT%20still%20have%20the)). Important in the context of moral counterpossibles and theistic ethics.

- **Sinnott-Armstrong, W. (2009).** In the same general debate as Morriston (2009). He suggests an alternative counterpossible: *If God commanded rape, then rape would still be wrong*, which he finds "plausible to most people" ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=match%20at%20L174%20false%20counterpossible,of%20any%20technical%20details%20about)). This strengthens the intuition that counterpossible conditionals can carry truth evaluable content (here, that even under the hypothesis of a divine command, a moral truth would remain). Part of a broader critique of DCT on modal grounds.

- **Choo, F. (2024). "A Non-Vacuist Response to the Counterpossible Terrible Commands Objection."** *Erkenntnis* (online 2024) ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=Critics%20of%20Divine%20Command%20Theory,that%20we%20lack%20reason%20to)). An open-access paper that engages the DCT debate by rejecting vacuism. Choo argues that one can deny the intuitive falsity of the counterpossible or otherwise defuse the objection without resorting to "it's all true anyway." He examines our intuitions and suggests they might not be as clear-cut or might be explained differently ([A Non-Vacuist Response to the Counterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=match%20at%20L58%20counterpossible%20terrible,by%20examining%20th

e%20underlying%20intuitions)) ([A Non-Vacuist Response to theCounterpossible Terrible Commands Objection | Erkenntnis

](https://link.springer.com/article/10.1007/s10670-024-00810-x#:~:text=counterpossible%20terrible%20commands%20objection%3A%20I,by%20examining%20the%20underlying%20intuitions)). Useful for seeing how a moral philosopher deals with impossible-world semantics in practice.

- **Enoch, D. (2011). *Taking Morality Seriously: A Defense of Robust Realism*.** (Oxford University Press). Enoch doesn't directly delve into impossible worlds, but in Ch.7 he discusses evolutionary debunking arguments and uses a modalized counterfactual perspective: e.g., if moral truths are necessary, certain counterfactual conditions for knowledge (like sensitivity) are automatically met (Microsoft Word - Five Kinds of Epistemic Arguments Against Robust Moral Realism for web). This has implications for the necessity of moral truth and how counterpossibles (though he doesn't use the term) might be leveraged in metaethical arguments.

- **Field, H. (1989). *Realism, Mathematics and Modality*.** (Blackwell). Field's work on the independence of math from logic gave the oft-cited example about the axiom of choice ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=changes%20the%20example%20to%20%E2%80%9CNothing,%28Field%201989%3A%20237%E2%80%938)). He uses a counterpossible (from the perspective of someone who thinks ZF+Choice is necessary) to argue about mathematical and logical necessity not coinciding. This is analogous to moral necessity vs logical necessity – a reminder that "necessary" can mean different things (logical, metaphysical, etc.), and impossible worlds can be relative to each notion.

These sources collectively cover the central themes of impossible worlds in modality and the specific issue of moral counterfactuals. They illustrate the development from early logical motivations to rich contemporary debates, providing both formal frameworks and philosophical analyses of

why we might need impossible worlds and what cautions or caveats come with them. Each reference above is tied to points discussed in the review, offering a trail for further reading on those specific aspects. The literature continues to grow as philosophers refine these ideas, making impossible worlds an exciting intersection of logic, metaphysics, language, and ethics. ([

Impossible Worlds (Stanford Encyclopedia of Philosophy)

](https://plato.stanford.edu/entries/impossible-worlds/#:~:text=characterising%20them%20as%20ways%20things,be%20an%20absolute%20impossibility%3B%20and))

References

[1] Leibniz's Modal Metaphysics - Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/entries/leibniz-modal/#:~:text=Philosophy%20plato,if%20he%20were%20not

[2] Impossible Worlds | Internet Encyclopedia of Philosophy: https://iep.utm.edu/impossible-worlds/#:~:text=1,the%20existence%20of%20impossible%20worlds

[3] Impossible Worlds | Internet Encyclopedia of Philosophy: https://iep.utm.edu/impossible-worlds/#:~:text=the%20existence%20of%20impossible%20worlds,worlds%20from%20being%20acceptable%20entities

[4] Impossible Worlds | Internet Encyclopedia of Philosophy: https://iep.utm.edu/impossible-worlds/#:~:text=To%20begin%20with%2C%20representing%20properties,Consider%20these%20two%20statements

[5] Impossible Worlds | Internet Encyclopedia of Philosophy: https://iep.utm.edu/impossible-worlds/#:~:text=,is%20not%20in%20Australia

[6] Impossible Worlds | Internet Encyclopedia of Philosophy: https://iep.utm.edu/impossible-worlds/#:~:text=disputable%2C%20view%20of%20propositions%20as,is%20false%2C%20and%20vice%20versa

[7] Impossible Worlds | Internet Encyclopedia of Philosophy: https://iep.utm.edu/impossible-worlds/#:~:text=fairly%20often%20hosts%20impossible%20events

[8] Williamson on Counterpossibles | Journal of Philosophical Logic: https://link.springer.com/article/10.1007/s10992-017-9446-x#:~:text=Logic%20link,2

[9] Impossible Worlds | Internet Encyclopedia of Philosophy: https://iep.utm.edu/impossible-worlds/#:~:text=,the%20time%20would%20have%20cared

[10] Impossible Worlds | Internet Encyclopedia of Philosophy: https://iep.utm.edu/impossible-worlds/#:~:text=have%20cared%20if%20the%20antecedent,worlds%20analysis%20is%20one%20candidate

[11] Impossible Worlds | Internet Encyclopedia of Philosophy: https://iep.utm.edu/impossible-worlds/#:~:text=Although%20intuitions%20are%20usually%20controversial,worlds%20analysis%20is%20one%20candidate

[12] Impossible Worlds | Internet Encyclopedia of Philosophy: https://iep.utm.edu/impossible-worlds/#:~:text=Whatever%20the%20machinery%20behind%20the,admit%20that%20they%20are

[13] Impossible Worlds | Internet Encyclopedia of Philosophy: https://iep.utm.edu/impossible-worlds/#:~:text=counterfactual%20conditionals%2C%20nonetheless,treat%20the%20putative%20triviality%20by

[14] Timothy Williamson, Counterpossibles - PhilPapers: https://philpapers.org/rec/WILC-37#:~:text=Timothy%20Williamson%2C%20Counterpossibles%20,vacuously%20true%20against%20recent%20criticisms

[15] Impossible Worlds | Internet Encyclopedia of Philosophy: https://iep.utm.edu/impossible-worlds/#:~:text=Disagreements%20in%20metaphysical%20disputes%20display,Similarly%2C%20proponents%20of%20modal%20realism

[16] Microsoft Word - Necessary Moral Truths and Theistic Metaethics.doc: https://philpapers.org/archive/DANNMT.pdf#:~:text=10%20An%20anonymous%20reviewer%20worried,indeed%20assuming%20that%20there%20are

[17] Microsoft Word - Necessary Moral Truths and Theistic Metaethics.doc: https://philpapers.org/archive/DANNMT.pdf#:~:text=metaphysically%20necessary%20truths%20of%20this,I%20am%20not%20sure%20that

[18] Microsoft Word - Five Kinds of Epistemic Arguments Against Robust Moral Realism for web: https://philarchive.org/archive/SCHFKO#:~:text=or%20safe,have%20had%20deeply%20

mistaken%20moral

[19] Formulating the Moral Argument | Reasonable Faith: https://www.reasonablefaith.org/writings/question-answer/formulating-the-moral-argument#:~:text=But%20these%20Nietztschean%20implications%20are,values%20and%20duties%20would%20exist

[20] Formulating the Moral Argument | Reasonable Faith: https://www.reasonablefaith.org/writings/question-answer/formulating-the-moral-argument#:~:text=cosmological%20argument%2C%20dependent%20on%20asserting,values%20and%20duties%20would%20exist

[21] Microsoft Word - Five Kinds of Epistemic Arguments Against Robust Moral Realism for web: https://philarchive.org/archive/SCHFKO#:~:text=understanding,and%20present%20communities%20with%20that

[22] Microsoft Word - Five Kinds of Epistemic Arguments Against Robust Moral Realism for web: https://philarchive.org/archive/SCHFKO#:~:text=beliefs%20can%20be%20on%20moral,2020%29%20for%20relevant%20discussion

[23] Impossible Worlds | Internet Encyclopedia of Philosophy: https://iep.utm.edu/impossible-worlds/#:~:text=The%20well,worlds%20insist%20that%20unless%20we

[24] Microsoft Word - Necessary Moral Truths and Theistic Metaethics.doc: https://philpapers.org/archive/DANNMT.pdf#:~:text=Second%2C%20I%20think%20the%20examples,Oxford%3A%20Oxford%20University%20Press

[25] Formulating the Moral Argument | Reasonable Faith: https://www.reasonablefaith.org/writings/question-answer/formulating-the-moral-argument#:~:text=2011,duties%20is%20that%20God%20exists

[26] Counterpossibles, consequence and context - Taylor and Francis: https://www.tandfonline.com/doi/full/10.1080/0020174X.2024.2326184#:~:text=that%20the%20

[27] Microsoft Word - Five Kinds of Epistemic Arguments Against Robust Moral Realism for web: https://philarchive.org/archive/SCHFKO#:~:text=Similarly%2C%20it%20is%20not%20obvious,there%20is%20a%20good%20causal