On the use of impossible worlds

MARTÍN SOTO, University of Barcelona, Barcelona, Spain

I study the use and epistemic standing of counterfactuals and impossible worlds. I argue for their fundamentally doxastic nature (thus collapsing or dissolving additional ontological inquiries), defend that there is no objective or privileged sharp delineation between possibility and impossibility, and postulate that the non-formalizability of impossible worlds is in fact their main source of usefulness. I connect these claims to Quine's Web of Belief, and more generally Rorty's Pan-relationalism.

INTRODUCTION

In this article I discuss impossible worlds. Let me sketch the pre-theoretic notion I'm trying to get at with that term. The *actual world* comprises all that is the case. For example, apparently contingent truths, like my name being Martín, as well as apparently less contingent truths, like 2+2=4. A *possible world* is a state of affairs for reality (or truth assignment for statements) which might differ in some way from the actual world, but whose difference doesn't seem extraneous, extreme, or too fundamentally alien: it is an easily conceivable difference, something contingent that could easily have played out differently.[1] For example, my being called David. Indeed, it doesn't seem like any fundamental law of nature is constraining my name to be Martín, this seems like a contingent truth that could have gone otherwise. An *impossible world* is also an alternate state of affairs differing from the actual world, but its difference does seem extreme, or non-contingent, or hard to conceive. For example, a world in which 2+2=5. There do seem to be pretty fundamental laws of mathematics constraining 2+2 to be 4, so we're left not exactly knowing how to interpret this alternate state of affairs. I will label as *counterfactuals* any non-actual world (thus, both possible and impossible).

In *Thinking the impossible*, Priest defends the theoretical status of impossible worlds. This can be understood as an argument from usefulness: Priest provides examples in which the logical formalization of some parts of math or philosophy naturally call for their use. And also from conceivability: We can theoretically entertain exotic claims like "A box being both empty and full", and use our existing theoretical toolkits or heuristics to study this hypothetical (even if we cannot visually imagine it, which is but one mode of thought).

In this piece, I will argue Priest is right. But I will do it from grounds pretty different from his, providing what I consider to be the "real philosophical reason" why Priest is right, or at least the most satisfactorily explanatory words I can write to that effect. These grounds are a collapse of counterfactuals as an epistemic phenomenon. I will also, in some places, and especially regarding Claim C below, argue Priest didn't go far enough. The concrete philosophical claims I will defend, that also structure the below sections, are the following:

A. Counterfactuals are a fundamentally doxastic phenomenon.
B. There's no privileged boundary between possibility and impossibility.
C. The unbounded generativeness of impossible worlds makes them both non-formalizable and useful.

After Section A, I also present, motivate and interpret Quine's concept of the Web of Belief, which is very relevant to Claim A, and epistemic collapses more generally.

Author's information: Martín Soto, martin.soto@longtermrisk.com, martin-soto.com.
[1]In particular, the actual world is a possible world, since none of its differences from the actual world (none) are extreme.

Throughout, argumentation style focuses on pumping intuitions, appealing to common usage and practicality, and generally emphasizes explanation and understanding of the phenomenon over water-tight philosophical grounding. This methodological choice is very related to the closing thoughts in Section B about pan-relationalism.

The stronger philosophical claim about ontological collapse was already defended by Butcharov in *Counterfactuals and antirealism*. But again, here I focus on presenting an intuitive and explanatory story about why that should be the case, and connecting these insights to the practical reality of conceptual use.

A. COUNTERFACTUALS ARE FUNDAMENTALLY DOXASTIC

Addressing Claim A, it's unclear what it means for a phenomenon to be fundamentally doxastic. I'll limit myself to the following practical operationalization in two claims:

- There is a robust doxastic justification for the conceptual, theoretical and pragmatic status of counterfactuals.
- There is no robust motivation from any other grounds.

For the first claim, note the conceptual use of counterfactuals lets us more efficiently navigate the world as physically or computationally bounded agents. For example, say I commit to the following: I will throw a fair coin, which will land inside a box where you can't see how it landed, and then offer you a 60/40 bet on Heads. The usual Expected Value justification for taking the bet builds on the subjective possibility (and indeed, equal probability) of your being in the Heads world, and your being in the Tails world. This first-person, epistemically uncertain point of view is all that matters for the purpose of conceptual usefulness, since that's the only position anyone ever finds herself in. In this situation, it would seem **counterfactuals emerge as subjective uncertainty about the actual**.

But is that characterization exhaustive? It seems not, as showcased by the following example. A scientist has performed many experiments on physical system A (a particular molecular configuration). She has thoroughly observed that, 5 seconds after system A comes into existence, it evolves into system B (a different configuration, the result of an energy-minimizing transition). Someone goes up to the scientist with a transparent, empty box, and asks: "If inside this box there was system A (instead of just air, as right now), what would be inside it in 5 seconds' time?". The scientist will answer "system B", and it intuitively seems like this is justified, and maybe even true (modulo, of course, the problem of induction). But the scientist is not really uncertain about the contents of the box. She is just playing along with the hypothetical. So our above narrow definition, which seemed to require literal uncertainty, cannot account for this case.

In the example, the scientist is using the same heuristic[2] she'd use in a literally uncertain situation, but now for a different purpose: sharing this heuristic with the questioner, showcasing how she reasons about the world. And what is, really, the information that's being shared then? It's a pattern in actual reality that the scientist has noticed. This exactly is the conceptual content of these counterfactual games (as opposed to their useful role, for example in uncertain prediction): they are **an**

---

[2]Here and below, by *heuristic* I mean any piece of conceptual apparatus employed for some pragmatic end, like making sense of past observations, predicting future observations, or pursuing one's goals in actual reality. My focus on them won't come as a surprise in light of the connections to Quine's Web of Belief, in which all conceptual objects are fundamentally a tool for the pragmatic goal of predicting observations.

efficient heuristic summary of patterns observed in actual reality[3]. Indeed, exactly the same pattern-summarizing happens when we say "If it had rained, the picnic would have been unpleasant". Or even when we ask a mathematician "What would have happened if Fermat's Last Theorem had been false?", and she answers "Elliptic curves of this weird shape would exist". Fermat's Last Theorem is true, and thus from assuming its negation 1=0 should follow (this basic and obvious heuristic she's aware of). But she correctly assumes that's not what we're really interested in, and instead we'd like to hear some of her more advanced heuristics, applicable to this mathematical branch in particular. We're interested in hearing explanatory heuristics, the ones that allowed her (or other mathematicians before her, in case she never was uncertain about such matters) to explore and investigate effectively the truth of these statements, even when still uncertain about them.

Or better said, these heuristic summaries already exist, and counterfactual reasoning is a way to either directly use them for predictions (when we seriously entertain the hypothetical with some probability), or to toy around with them, recombining and expanding their conceptual structure and interdependencies, to obtain deeper, more general or better heuristics[4].
As an exemplification, the scientist could learn the heuristic "A -> B" (a pattern in actual reality) even if she never considers counterfactuals. But then, if faced with true uncertainty about whether system A is inside the box, she will be able to predict observations correctly, by acknowledging she might be in any of the two "counterfactuals", and playing along with the hypothetical as if it was the case. But furthermore, she will be able to recombine this heuristic with others in her conceptual toolkit, so as to derive some more general principle ("all systems with property X evolve into system B"), or notice relevant internal structure ("it seems it's this part of the molecular configuration that drives the change").

This whole doxastic explanation can be understood as a pragmatic argument from conceptual usefulness. But very similarly, it can also be interpreted as explaining the linguistic use of counterfactuals, or providing a historical account for their conceptual existence and saliency, or their theoretical status. Which, according to the linguistic turn or similar ontological collapses, would leave nothing else to be explained (and indeed, see the next section for this connection).

For the second claim ("no robust motivation from any other grounds"), I'd like to start with a disambiguation. One might note the concept of alternate possible universes is employed in certain physical theories (or their interpretations) as objectively and independently existing parts of reality, instead of simply an epistemic heuristic. And here I will say this is a different use of the word "possible world" than the one I'm studying, because such theories or interpretations treat these possible universes as part of actual reality, whilst my pre-theoretic definition of counterfactuals concerns exactly that which falls outside of this realm. More concretely, in such cases I will identify "actual reality" with "the whole multi-verse", instead of "a single possible universe".

---

[3]Of course, it wouldn't even make sense to talk of patterns observed in counterfactual reality, since counterfactual reality is, by definition, the unobserved.
[4]Said recombination, of course, is itself mediated through previous or more general heuristics. Indeed, in a view like the Web of Belief (employed below), it will be heuristics all the way down. And also, it will be existing Webs which determine how to vary and compare Webs.

And then, of course, our whole discussion of possible worlds (now "possible multi-verses") replicates. Interesting questions arise about what marks this delineation between actual and non-actual[5].

As a more general argument, given our above characterization of counterfactuals, it's hopefully become clear why their mind-independent or ontological standing is irrelevant to understanding their use. Indeed, being nothing more than conceptual manipulations of patterns in actual reality, nothing outside of (or far and above) actual reality will shape and explain them[6]. One could even take the radical (but natural) next step, and claim this shows the philosophical question of their mind-independent or ontological standing is meaningless, or not useful (again, in line with the linguistic turn or ontological collapse).

In any event, through this interpretation we have confined counterfactuals to the doxastic realm, and apparently lost none of their content or purpose. While it's always hard to thoroughly argue the negation of an existential, this analysis seems to point towards the doxastic one being the minimal and natural explanation of counterfactuals, such that any other one will be routing through it, doing the same work with more ontological baggage, or anyhow re-wordable in doxastic terms (although this might be a property not particular to counterfactuals, if we go far enough in our ontological collapses, as explored in the next section).

QUINE'S WEB OF BELIEF

So we've seen these heuristic summaries (generalized induction) have two distinct use cases. On the one hand, they can be directly applied to predicting actuality. Although this is only possible when we are in certain doxastic positions. For example, the scientist needs to believe there is some positive probability of a system A existing in the future. Otherwise, that particular heuristic won't ever become applicable. On the other hand, and even when the heuristic is not directly applicable, it can inform our (or others') different heuristics (that might themselves be applicable). Of course, there is a sense in which the use of any such heuristic bottoms out in predicting the actuality. And indeed, if we were to judge all of our particular heuristics as a single undifferentiated jumble, we could understand this as a simple, very complex heuristic. This makes sense, since the boundaries between them are not sharp, and indeed that's what Quine's Web of Belief story would emphasize.

Many of the above arguments and intuitions would seem trivially true if we already bought wholesale the Web of Belief story that I now present. And indeed, I think this more radical collapse of ontology is the "real philosophical reason" why Priest is right, or at least the most satisfactorily explanatory words I can write to that effect. Although certainly buying this story wholesale is not necessary to appreciate the above arguments.

In *Two dogmas of empiricism*, Quine argues against the existence of a sharp, well-defined distinction between analytic facts (true purely by virtue of their form or meaning) and synthetic facts (requiring additional, empirical information to establish). Although I won't replicate this debate here, Quine

---

[5]Getting ahead of myself, under the Web of Belief story it will turn out the physicist's "possible universes" are conceptual parts of the Web of Belief, while my "counterfactuals" are variations on the existing Web of Belief (including the null variation, for the actual world). But going even one step further, given the impossibility of giving a closed form for all possible Webs (see Section C), and the indelible partiality of pan-relationalism, this distinction (being inside the Web vs not) will also be a fuzzy one.

[6]An even cruder argument (more obvious, and less explanatory with regards to the structure and workings of counterfactuals) would be to say: all our conceptual and theoretical use of counterfactuals happens within actual reality, so nothing more is needed to understand them.

argues that proposed definitions of analyticity lead to either circularity or infinite regress. And furthermore, attempted reductions of purely empirical evidence to logical constructs over our sense data have not just pragmatically failed, but also seem circular in principle.[7]

Instead, Quine paints a picture in which our whole epistemic theoretical edifice is a tangled web of conceptual beliefs, models and structures, in which traditionally analytic topics (like logic) and traditionally synthetic topics (like physics) are not neatly compartmentalized. Their only difference is relational and non-absolute, relating to which role they play in the theoretical edifice. Some basic truths of logic might be very robustly connected to everything, and very close to the center of the web, meaning it's a theoretical model very central to our understanding of the world, and we would need a lot of radically new and unexpected observations (whose shape we might not even be able to fruitfully imagine as of now) to alter it. On the contrary, some more contingent truths of physics, like the weather today, or even the standard model of subatomic particles, are less robust and closer to the periphery, playing a useful role but possibly being altered by just a few new and unsurprising observations. In this picture, a human receives observations[5], and slowly builds a Web of Belief whose sole purpose is to better predict future observations, maybe by coming up with very complex hypotheses (like the laws of physics) for patterns in or generators of the stream of observations.

There are a few things to like about this picture. First off, it seems close to what the brain really does, which helps characterize and explain real phenomena. While of course the brain doesn't have annotated conceptual nodes, its information encoded in decentralized complex neural firings and connections seems like a close analogue. Similarly illuminating analogues exist with Machine Learning architectures, or the predictive processing literature.

Secondly, it can help us think effectively about modelling the world. Here's one example. We notice the relationship between a stream of past observations and Webs of Beliefs (or world-models) explaining them is one-to-many[8]. For instance, I might posit the gravitational constant has the value physicists claim, or I might alternatively posit its value is actually higher, but a unicorn from another dimension is constantly laboring to bring it down, with perfect precision. Both hypotheses will predict the same observations. Thus, we need some additional criterion to choose between Webs of Beliefs. This could be for example aesthetic ("I like unicorns in my theories"), or also pragmatic ("I want my model to be as computationally frugal as possible"). Of course, this is a classically well-known state of affairs in philosophy of science, which has traditionally been addressed by invoking Occam's razor, a simplicity prior over hypotheses. And the Web of Belief / world-model framework represents it natively.

Finally, and as shouldn't be surprising (since the Web of Belief crudely collapses everything to epistemics), it also helps formalize the dissolution of some metaphysical conundrums. Here's an example. To the question "Does God exist?"[9], Wittgenstein answers not Yes or No, but "The question

---

[7]A more modern epistemic-centric account, derived from Jeffrey's *Radical Probabilism* ("no facts are ever known absolutely for certain"), explains even more fundamentally why such a reduction cannot work in philosophical principle: because there also isn't a clearly delineated notion of "observation" or "sense data", sharply separated from other internal conceptual derivations. One such account is *cybernetics*, characterizing both minds and conceptual structures as a tangled mess of wetware information processing (indeed, very close to Quine's Web of Belief). It will similarly note that there is no clear point at which "an observation happens" or "is received", and instead a fuzzy boundary around the wetware, without a clear divide between external (empirical, synthetic) data acquisition and internal (theoretical, analytic) processing (see Demski).

[8]Similarly to how the relationship between strings of bits and computer programs generating them is one-to-many.

[9]Of course, if by God we meant anything with observable effects (a bearded man coming down from the sky), this would simply become a scientific, world-modelling question, dealt with just as we deal with predicting the trajectory of a ball. For our purposes, I mean a God, an platonic essence, or a *res cogitans*, that doesn't have any bearing whatsoever on observations.

is non-sensical, or meaningless". While Wittgenstein seems to use different philosophical tools, we can naturally support this from the Web of Belief framework. Given any set of observations, there are two different Webs which perfectly explain them: one doesn't contain God (and just talks about physics, biology, etc.), and the other is exactly identical but additionally contains a node labeled God. This node is not connected to anything, or at least anything that affects observation prediction, since indeed by definition it has no bearing on such matters. It's thus clear, as in the paragraph above, that it's a matter of under-determined doxastic choice whether I choose one or the other. Occam's razor would recommend the God-less model. Some different criteria for model disambiguation would recommend otherwise. But it's clear this question has become arbitrary and meaningless, a free parameter in our model-making.

Note in the above I identify existence with being a part of my conceptual Web of Belief. That seems indeed, by definition, like the correct criteria for conceptual existence. But we usually mean something different by the term, and indeed while both electrons and the number 17 are represented in my Web, if I'm a physicalist I'll think electrons have a different existence status. Indeed, it will be useful for many purposes to have a (non-absolute, fuzzy) distinction between physical and conceptual existence. This position will simply state that the only difference between parts of my Web with and without the "physical existence" label is the role they play in my model and predictions.

As regards to counterfactuals, since we have characterized inductive heuristics as the Web of Belief, we now identify counterfactuals (the toying around with these heuristics) as variations on our Web of Belief. For example, we might change the truth value of a node ("the weather outside"), and see which consequences this has throughout our Web, that is, how our heuristics generalize to this alternate configuration.[10]

To this presentation of the Web of Belief, and also to my previous less radical arguments for the doxastic nature of counterfactuals, one can counter-argue my assessments are already theory-laden. I am already using theory, or pre-theoretical intuitions or conceptions, to make my point. For example, in explaining the Web of Belief I invoked concepts like "observations", "conceptual", "prediction", and even in my previous arguments I used concepts like "physics", "pattern", "useful". But this might seem unsatisfactory in both cases. For the Web of Belief, one might say

> you are using concepts from your existing Web of Belief to present and defend the concept of the Web of Belief itself[11], but exactly because of that, your notion or concept of it might be too narrow (not encapsulating all possible shapes of Web), or unconvincing to others with different Webs.

And for my previous arguments about counterfactuals, one might say

> you are using certain pre-theoretic intuitions or deliberations to discuss the structure of counterfactuals, but what about the (granted, very wacky) counterfactuals in

---

[10]Of course, if we try to more concretely pin down the structure of our Web, we notice these counterfactual assessments (variations on the Web) need to be grounded in, or carried out from, the Web itself (our conceptual space). Or we could have a hierarchical model, more natively representing how we use some concepts to vary others, but surely we will in some contexts need to refer to its whole, and non-formalizable (or at least non-static) self-referentiality will be necessary.

[11]Indeed, one could see a kind of Wittgensteinian "impossible linguistic self-reference" going on, when from within my frame (my Web, my theory, my concepts, my way of thinking, my language) I am pointing at the edges of said frame, trying to help you notice them. A feat which is never literally possible, since the frame is as invisible as the frame of our field of vision, since a language cannot discuss its own limits, it can only obtusely hint at some general understanding of them.

> which even these basic intuitions are different? For example, what about a counterfactual (maybe impossible) world in which the concept of "pattern" doesn't work the same way as in this our world? Your explanations only seem useful or sensical once you have such minimal concepts available, and thus you are not truly characterizing the nature of counterfactuals in general, just the theory available within some of them.

To this, I can only respond that the objection is an insurmountable limitation of language and philosophy. The *Münchhausen Trilemma* states all argumentation either bottoms out in some axioms, is circular, or leads to infinite regress. Here, I'm just using some properties of the pre-theoretic concepts as axioms. This will always be a fundamental limitation when discussing concepts or argumentation itself: we will always be able to note that we haven't addressed the matter in full generality, since concepts or argumentation can always conceivably exist outside of these particular axioms. In fact, this fundamental realization is very related to the nature and use of impossible worlds, as discussed in Section C below.

B. POSSIBILITY IS NOT SHARPLY DELINEATED

I will argue for Claim B ("no privileged boundary between possibility and impossibility") from four different, partly overlapping perspectives.

To start with, the doxastic motivation from usefulness remains just as strong when we face uncertainty or counterfactuals traditionally labelled as "impossible" (for example, logical). Say, for example, you face a bet on the parity of a far-off digit of pi, instead of the landing of an empirical coin. One way this could play out, is I ask you for a big number before you know which game we'll play (so that you have a guarantee I haven't adversarially chosen the digit of pi), and then I give you 5 seconds to decide whether to take a 60/40 bet on said digit of pi being even. Since we have a strong (yet unproven) mathematical conjecture that far-off digits of pi are pseudorandomly distributed, you seem to have theoretical reason to take the bet. Just as before, your reasoning is naturally represented as being uncertain about which of the doxastically possible "counterfactuals" corresponds to actual reality. Just now the counterfactual is with respect to a mathematical, instead of an empirical fact.[12]

One might worry I have been too quick, and a sharp distinction can still be drawn between both situations. For example, it seems vital for this example that you don't have unlimited time to think (or a calculator available). And so we might be tempted to define impossible counterfactuals as "those that, were you to have an unbounded amount of time or compute available, would appear clearly not actual". But again I will argue such a definition is not absolute. In fact, it is circular on your notion of counterfactuals. How exactly is the unbounded amount of time implemented in the situation? For example, does the rest of the world freeze?[13] This is already banking on a fixed way to take counterfactuals. Will this always be a valid counterfactual move? Or similarly, how is the compute wired up to your brain, or used by you? Isn't observing a calculator an empirical observation? Which is the threshold for appearing "clearly not actual"?[14] Is there a fundamental difference, from your epistemic position, between thinking with your eyes closed, thinking with your eyes open, using a calculator, and lifting a box to see what's inside? The next paragraph

---

[12]And in case we worry about a distinction between mathematical and logical facts, we can exactly equivalently make the example fact logical (the satisfiability of a logical sentence).

[13]We can even construct more exotic decision-theoretic scenarios in which the bettor doesn't even want to perform the computation, as much time and compute as it might have, see Soto.

[14]In fact, can we ever be certain in such a way, as per Jeffrey's *Radical Probabilism*?

defends further the unavoidable graduality and partiality of these distinctions.

Secondly, and again using "logical" as the best candidate stand-in for "impossible to vary", there is no sharp distinction between logical and non-logical observations. Even in the usual example of throwing a coin, it is conceivable that with the visual observations in the moments prior to throwing the coin, we had enough information to perfectly compute which side it would land (if we think physics is deterministic enough). We don't know whether this constitutes purely logical, purely empirical, or mixed uncertainty.[15]
What about using a calculator? That seems like an empirical observation, but is it providing logical information? What about working out half of the computation in your head, and the other half in a calculator? What about doing it by hand? The graduality of the distinction, and the non-existence of a privileged cutting point, seems intuitively evident.[16]

Thirdly, the possible/impossible distinction seems equivalent to the synthetic/analytic distinction, and thus they are both non-sharp and arbitrary because of the same reason (if we agree with Quine). Indeed, a natural, sometimes implicit, delineation of possible/impossible is for possible worlds to differ from actuality only in synthetic facts, while impossible worlds also differ in analytic facts. So it's natural, if the synthetic/analytic sharp distinction to fall, to expect the other one to fall through. I won't go as far as to claim this is a guaranteed implication, since it is conceivable (albeit less intuitive) to posit a privileged, non-relative definition of possibility that doesn't route through the definition of analyticity.

Finally, according to the Web of Belief picture, variations that correspond to possible and impossible arrangements will be doxastically relative to that same Web of Belief, and need not neatly compartmentalize into a binary[17]. Indeed, there is no fundamental difference between varying a node holding the truth value of a mathematical statement, or that of an empirical observation (were we to represent Webs as causal graphs). Just a quantitative and partial difference due to their different positions and relations. Similarly, there might seem to be a sharp difference between changing the value of a node, and changing the structure of the causal graph itself, or even more radically representing the Web in a different mathematical structure. But, as discussed below in Section C, we can't have a closed formal specification of the Webs' shape, or the meta-theory or language in which they are constructed (or better said, doing so would imply already making some arbitrary choices about counterfactuals or possibility, and losing out on the conceptual benefits of other arrangements). So this distinction will also be relative to a doxastic position (for example, the

---

[15]Of course, we could simply appeal to a perfect quantum coin (were physics found to truly work that way), but the same arguments about the circularity of analyticity will apply. Or, closer to the spirit of this example: you're constantly uncertain about the simplicity and properties of infinitely many matheamtical structures that could define a theory of physics together with the (maybe pre-determined) behavior of its quantum coins. So, depending on your prior probabilities on these different hypotheses (maybe weighed by simplicity), it's always possible that observing the outcome of this quantum coin gives you information (moving probabilities around) about whether this or that (hard to compute) mathematical structure produces the outcome you just observed, or about its simplicity.

[16]One might hope we'd be safe by drawing the sharp distinction at "working it out exclusively in your head, independently of external stimuli relevant to the computation". But then, what if you hear an external sound that (maybe unbeknownst to you) has a 0.01 correlation with the truth of the mathematical statement, according to your current world-model? And even, what exactly constitutes a computation going on exclusively within your head? (This, again, is a known lesson from cybernetics.) What if some externally generated gravitational pull affects slightly the behavior of some neurons? Again we see settling on a sharp definition here is equivalent to settling on a (to some extent) arbitrarily chosen notion of possibility (whether gravitational waves are varied across possible worlds, etc.).

[17]And this is actually little more than a restatement of the previous paragraph from a different perspective

shapes of Webs that we actually have or have not conceived of up until now).

Yet to all the above, one might counter that some statements are so self-evident that they surely must count as impossible to vary (or equivalently, analytic). For example, a world in which 2=1 should certainly, by all lights, and by the opinion of everyone involved, be an impossible world. To this intuitive point I present two answers.

First, and as discussed above, even when we assign probability 0, we can nonetheless play along with the hypothetical, and extract some conclusions like 1+1=2+1 (without needing to resort to logical explosion). And do we observe some fundamental difference between this playing along, and the one we do when entertaining a hypothetical we are actually uncertain about, for direct prediction purposes? Different (possibly overlapping) heuristics are employed, certainly. And they have different ripple effects across our beliefs. But it seems like these differences are more quantitative than qualitative, and there is no fundamental difference between these two epistemic processes.

Second, and more radically, I can again highlight the fundamental conceptual relativity of it all, even of 2=1. Certainly to you and me the likelihood of 2=1 being the case seems 0,[18] but who knows what distinct minds (or better yet, conceptual structures and models of the world) can look like?[19]

Actually, this last reasoning overlooks an important subtlety. What do I mean by a Web different from mine believing 2=1? Am I content with any old web one of whose nodes is labeled "2=1"? Or do I need some apparently fundamental structural features of "2", "1" and "=" to count it as mirroring the statement in my Web, to ensure we're talking about the same thing? After all, it is true that if I impose strong enough such structural features, I will count no Web as believing 2=1. Suffice it to say that the user of the Web will have to doxastically assess that herself (just as she assesses questions about meaning and other Webs), and as always in an uncertain, dynamic and vague manner. So first off, due to her not being omniscient, she might be surprised to discover sensible ever-more-exotic Webs in which a node satisfies these properties of the number 2 (that she deemed central), but not those other ones. And even more: for any structural features she might impose on "2", "1" and "=", they will themselves be part of her mathematical beliefs, and she will be able to conceive of these going a different way (for example, so that the structural features don't make it impossible for another Web to believe 2=1).

This is indeed an important property of Webs that Quine highlighted: holism about the theoretical edifice, meaning it can only be evaluated as a whole (and from within), and some parts of mine need not neatly correspond to other parts of yours. This is an instance of *Pan-relationalism*: the elimination of essences, such that only relations to the rest of the theoretical edifice characterize the object. Which, according to Rorty, is a position simultaneously but independently arrived at by analytic philosophers like Quine, and continental philosophers like Derrida (albeit in very different languages, and albeit the latter would probably want to jump out of this classification).

Some of my dissolution of boundaries above is very much in the pan-relationalist spirit. And as explained by Rorty, if I were to push this direction to its conceptual conclusion, I might need

---

[18]Although again, by *Radical Probabilism*, or more naturally by how the brain works, it might not be literally 0.

[19]It has been proposed, for example, to use some mathematical proxy, like minimal proof length, to assess the "level of self-evidentiality" of statements (see Jago, 2007). But any such proxy will deviate from "actual evidentiality to me" in some scenarios (except for trivial self-referential proxies), and in any event the choice of such a metric will itself be a part of our theoretical edifice that we should be able to vary. So it seems philosophically more natural to simply accept this relativity.

to accept I cannot assert conclusively and without reservation these claims about the structure of counterfactuals, as distinct from other possible claims, and instead can only gesture at such conceptual musings being sensible or useful from certain shapes of theoretical edifice. That it would be a mistake to understand this explanation as an objective epistemic judgement, since on the contrary the bottomline is all such judgements are relative. Indeed, this is the same fundamental limitation I noted at the end of Section A.

C. IMPOSSIBILITY IS UNBOUNDEDLY GENERATIVE

To address Claim C ("impossible worlds are non-formalizable"), I will start with an example.

Even if intuitionist logic were to be incorrect (whatever that means), Priest argues for the inclusion of (im)possible worlds in which it is correct. Otherwise, the two following implications would be equally vacuously true:

– If intuitionist logic is true, then the Law of Excluded Middle fails.
– If intuitionist logic is true, then the Law of Non-Contradiction fails.

But, intuitively, we want the first one to be true in a way that the second isn't. He thus enunciates the general principle, that *everything must hold at some world* (including false things, like intuitionist logic being true).

But then, why not go one step further? Why not have a world in which intuitionist logic is true and LEM doesn't fail? This is, after all, conceivably a claim, and it might hold the place of antecedent in some similarly interesting implication. But then it would seem Priest's original implication wouldn't remain as true, due to having a counterexample (im)possible world. Maybe we want different tiers of truth or implication to account for these intuitions.

Depending on how we have arbitrarily, we might disregard sentences like "intuitionist logic is true", or its conjunction with "LEM fails", as outside our language. But of course, this is again an arbitrary doxastic choice, since the structure of language itself is part of our conceptual edifice.

But why would fixing this structure, and treating it as objective guardrails, be unwarranted or negative? One reason is that we cannot say a priori which counterfactuals or possibilities will be useful or actual, and so treating structure as fixed would possibly keep us from important conceptual advances. As a topical example, non-Euclidean geometries were for some time considered impossible, or contrary to analytic reason, but it was later discovered physical reality was better described by such geometries. Indeed, the story of science has many times faced an expansion towards the most exotic direction of impossibility (from the doxastic position at previous times). Priest postulates that something similar has happened in recent decades in the Philosophy of Logic, where ever more exotic and expansive theoretical objects or philosophical possibilities have turned out conceivable and useful.

In summary, impossible worlds are an unceasing source of previously inconceivable considerations. Their historic usefulness comes exactly from clashing with our pre-existing formalisms, and thus forcing their expansion and generalization. Trying to fit them within a fixed formalism would be both unwarranted and arbitrary, and neglecting their whole purpose.

REFERENCES

[1] Butcharov. *Counterfactuals and antirealism.* In *South Asian Research Journal of Humanities and Social Sciences*, Volume 1, Issue 1, 2019.
[2] Demski. *Radical Probabilism.* In *LessWrong Forum*, 2020.
[3] Jago. *Hintikka and Cresswell on logical omniscience.* In *Logic and Logical Philosophy*, Volume 15, Issue 4, 2007.

[4] Jago. *Impossible worlds*. In *Noûs*, 47, 2013.

[5] Jeffrey. *Epistemology Probabilized*. 2002.

[6] Priest. *Thinking the impossible*. In *Argumenta*, 2, 2, 2017.

[7] Priest. *What is a non-normal world?*. In *Logique & Analyse*, 139-140, 1992.

[8] Quine. *Two dogmas of empiricism*. 1951.

[9] Rorty. *Pan-relationalism*. In *Pragmatism as Anti-authoritarianism*, 2021.

[10] Soto. *Logically Uncertain Decision-Making*. In *PIBBSS Symposium*, 2023.