

# **Modal Rationalism**

*By Richard Chappell*

## Introduction

How can we know about what is possible or necessary? Empirical investigation might inform us of how the world is. But how can we learn about what isn't, yet *could have been*, the case? How are we to distinguish such contingent falsehoods from genuine impossibilities, which are not only false but necessarily so, such that there is no way that they could have been true? What, we may ask, is the scope of possibility?

We may clarify these problems by recasting them in the idiom of “possible worlds”. The *actual world* corresponds to the universe, or all that actually exists. It is the way things actually happen to be. But things could have turned out differently: there are other *possible worlds*, or ways a world could be. It happens that there are not any unicorns. But presumably there could have been. That is, there are possible (non-actual) worlds where unicorns exist. More generally, we can identify *possibility* with truth in some possible world, and *necessity* with truth in all possible worlds. A statement is *contingent* if it is true in some possible worlds but not others. In this framework, *modal space* – the range of possible worlds – provides the scope of possibility.

Often we are interested only in a restricted subset of modal space. For instance, when thinking about what's *nomologically* possible – possible given the actual laws of nature – we should only consider those possible worlds that have the same laws of nature as our own. But perhaps these laws could have been different, so that there are possible worlds where this is the case. Such

worlds would not be “nomologically possible”, of course, but they would still be possible in a more fundamental sense, which we can call “metaphysical possibility”. They are ways the world could have been. It is this broad notion that the present work is concerned with.

Sometimes we use talk of possibility in a subjective epistemic sense, merely to highlight what might be true for all we know. To a novice mathematician, it might seem “possible” in this sense that there is a greatest prime number. But though he might not yet know it, others have proven that this cannot be true. It is not *metaphysically* possible: there is no way the world could have turned out that would make his claim true. Rather, it is false in all possible worlds that there is a greatest prime number. As this example illustrates, metaphysical possibility is an objective notion, concerned with how the world could have turned out in fact, rather than how it might be for all we know. We should similarly distinguish metaphysical necessity from epistemic certainty. For instance: I certainly know that, as it happens, I exist. But although my existence is (epistemically) certain, I presume it is not necessary. Things could have been different – there are possible worlds where I don’t exist. It’s just that those worlds are incompatible with the contingent evidence that is available to me through introspection, so that I can know them to be non-actual. That doesn’t make them any less possible in the metaphysical sense.

A more promising epistemic distinction holds between *a priori* and *a posteriori* statements. This concerns the mode or basis of justification, rather than strength of credence. Let us say that a statement is knowable *a priori* iff<sup>1</sup> it can be conclusively justified by reason alone, without appeal to experience.

(Experience may be required in order to acquire the concepts one is reasoning about, but that need not cast doubt on the possibility of a priori *justification*. Experience may be a necessary precondition for thought without thereby playing a justificatory role.)<sup>2</sup> Truths of mathematics and logic are paradigm examples of the a priori. These may be contrasted with empirical or *a posteriori* statements, which require experiential justification. Our familiar everyday beliefs about the world tend to be of this latter sort, as are the claims of empirical science.

It is natural to expect a close connection between this epistemic distinction and the metaphysical distinction between necessity and contingency. After all, contingent claims are true in some worlds but not others, so how could one know them to be true without first looking to see *which* world one is in? This line of thought suggests that in order to know something a priori, without appeal to experience, it must be true in all possible worlds, i.e. necessarily true. Conversely, if a claim will turn out true no matter how the world is, then looking at the world might seem superfluous. If it is necessary, then it should be knowable a priori. Combining these claims yields the Coincidence Thesis:

**(CT):** For any statement  $S$ :  $S$  is a priori iff  $S$  is necessary.

CT offers one way to establish **Modal Rationalism**: the thesis that we have a priori access to the space of possible worlds. On this view, we can discover what's possible by determining what is consistent or cannot be ruled out a priori. Empirical investigation comes later, when we wish to see which possibility is actual. Experience may then inform us of how the world *is*, but

we do not need it in order to learn how the world *could* have been. This prior task may be conducted by reason alone.

Remarkably, Kripke established that CT is in fact false. We can provide counterexamples both in the form of contingent a priori statements, and others that are necessary yet a posteriori. One may be tempted to conclude that empirical inquiry is as indispensable for learning about *other* possible worlds as it is for learning about the actual one. On this view, even claims that are a priori coherent might later be shown by science to be strictly impossible, i.e. false in all possible worlds. This challenges the modal rationalist's claim that we can (in principle) grasp modal space through reason alone.

It's worth noting that this dispute has broader implications for philosophical methodology, by influencing the kinds of arguments that metaphysicians may employ. Several important arguments depend on the inference from conceivability (or a priori coherence) to possibility legitimated by modal rationalism. For example, Chalmers (1996) argues from the conceivability of "zombies" – creatures physically identical to us but lacking phenomenally conscious experiences – to their metaphysical possibility, and hence the falsity of physicalism. That is, if modal rationalism is correct, then conceivability arguments of this sort can establish that the physical alone does not suffice to guarantee consciousness. Or consider the following variation of Moore's "Open Question Argument" against meta-ethical naturalism: assuming that the coincidence of *goodness* with any natural property (say *happiness*) would remain an "open question" even after ideal a priori reflection, modal rationalists may infer that the two could therefore

come apart, and hence are distinct properties. More generally, given the necessity of identity, modal rationalism enables the powerful argumentative strategy of inferring actual non-identity from merely conceivable non-identity. Without it, such inferences would become suspect; and the philosopher's toolkit would be that much more limited.

This thesis examines whether modal rationalism is defensible. Chapter One explores Kripke's arguments, and shows how the semantic framework of Two-Dimensionalism promises to restore the link between reason and modality that found naïve expression in CT. Chapter Two introduces the idea of "strong necessities" that elude even the 2-D framework. Finally, the third chapter explores a metaphysically "realist" conception of primitive modality that can help motivate the radical challenge of strong necessities, though I suggest a way to reconcile realist modal primitivism with modal rationalism.

## Chapter One

### §1.1 Kripkean Semantics and the Contingent *a priori*

I am thinking of a particular person X. He is very well known, and there are numerous expressions in our language – some quite colourful – that could be used to refer to him. For a drab example, I can pick him out by definite description, as ‘the President of the United States’, or else by the name ‘GWB’. These terms are co-extensional: both pick out the same entity X, namely, Bush. But the two expressions differ in modal character. For consider that possible world  $w$  where Kerry was elected president instead of Bush. When we apply our terms to that world, we find that their referents diverge. ‘The President of the United States’ picks out Kerry in world  $w$ , even though it refers to Bush in our actual world @. Kripke argues that a proper name such as ‘GWB’ refers to the same individual X – Bush – in both @ and  $w$ , and indeed in all possible worlds where he exists.<sup>3</sup> A term that picks out the same object in all possible worlds may be called a *rigid designator*. Names, such as ‘GWB’, and natural kind terms, such as ‘water’, are paradigm instances of rigid designators. They stand in contrast to non-rigid descriptions, such as ‘the President of the United States’, which may be satisfied by different individuals in different worlds.

The content of a rigid designator may depend on how the world actually turns out, as when contingent descriptions are used to *fix the reference* of a rigid designator. Whatever object happens to actually fit the description, *that very object* is then denoted by the rigid designator in all possible worlds – even

those in which the object no longer satisfies the original description. Hence the reference-fixing description and the rigidly designating term might come apart in non-actual possible worlds. In such a case, we can know *a priori* that the designated object meets the description, even though it is metaphysically contingent. For instance, we might use a particular stick *S* to define (fix the reference of) the length ‘one metre’. But once this length is so fixed, we can imagine possible worlds where *S* is a different length, longer or shorter than its actual length of one metre. ‘The length of *S*’ and ‘one metre’ would no longer be co-extensional in such worlds, assuming that the latter is a rigid designator. Hence, Kripke argues that ‘*S* is one metre long’ is a contingent yet *a priori* statement.<sup>4</sup> We don’t need to look at the world in order to learn that it’s true, because however long it actually turns out, that is also how long one metre is. Nevertheless, when we consider a *counterfactual* possible world, we hold the actual world fixed. That is, when applying the rigid designator ‘one metre’ to another possible world *w*, it still refers to the *actual* length of *S*, rather than to the length of *S* in *w*. It does not vary from world to world the way its associated reference-fixing description (‘the length of *S*’) does.

Of course, Kripke allows that the people in *w* might use the expression ‘one metre’ to denote the length of *S* in *w*. So they could speak truly in uttering the sentence ‘*S* is one metre long.’ But that is because the term ‘metre’ means something different in their mouths. We are concerned with what *our* terms mean when applied to counterfactual worlds.<sup>5</sup> This must not be confused with what sentences would be true in the counterfactual world’s language. (There is a possible language in which ‘two’ means five. That does not make our sentence ‘two plus two equals ten’ possibly true in the relevant sense.) In *our*

language, we may stipulate, the term ‘one metre’ rigidly designates the actual length of *S*. So, in our language, ‘*S* is one metre long’ is false of a world *w* where *S* is longer or shorter than it is in our world.

To clarify this point, consider the statement:

(A) “*S* is as long as *S* actually is”

While (A) is presumably *a priori*, it may yet be false in *other* possible worlds, so long as the “actually” operator is understood as fixedly referring to *our* world. There are worlds where *S* is a different length from what it actually is, after all. So statements like (A) are contingent despite being *a priori*.

To make this more precise, let us understand the modal operator ‘actually’ as signifying ‘in @’, where ‘@’ rigidly designates the actual world.<sup>6</sup> Consider some contingent proposition *P* that is actually true. Because *P* is contingent, there will be some possible world *w* at which it is false. At such a world, *P* is false, despite it being the case that *P* is true at @, i.e. ‘actually *P*’ is true. Because ‘*P*’ and ‘actually *P*’ differ in truth-value at world *w*, we find that ‘*P* iff actually *P*’ is false at *w*, and thus not metaphysically necessary. But again, ‘*P* iff actually *P*’ is presumably *a priori*: we don’t need to look at our world in order to know that the things true in it are actually true. It follows trivially from the definition of ‘actually’. This semantic trick makes clear how there could be contingent *a priori* statements. It is trivial *a priori* that ours is the actual world, even though there are other possible worlds that aren’t. We can get similar results by rigidifying other indexicals: “I am here now (if I exist)” is presumably *a priori* – I don’t first need to check *where* I am before knowing it to be true – but not metaphysically necessary, since I could have

been somewhere else. The key is that the referent of ‘here’ will be fixed by wherever I actually am. In this sense, rigid designators *epistemically* co-vary with their reference-fixing descriptions, hence making the coincidence a priori knowable, despite being metaphysically fixed once we shift our attention to counterfactual worlds.

### §1.2 *A posteriori necessities*

Rigid designators pick out the same entity in all possible worlds. So any two such designators that are actually co-referential will thereby also co-vary across every other possible world. Given that Cicero *is* Tully, so that our names ‘Cicero’ and ‘Tully’ both rigidly designate one and the same person, it follows that *necessarily*, Cicero is Tully. There is no possible world for which our two names pick out distinct individuals. But the sentence ‘Cicero is Tully’ is certainly not knowable a priori. The thought it expresses – and so, derivatively, the sentence itself – is an *a posteriori* necessity. It is conceivable (in the sense of not being a priori incoherent), but metaphysically impossible, that Cicero and Tully are distinct.<sup>7</sup>

Aside from identity statements, Kripke also argued that a posteriori necessities could be found in claims of natural kind membership, where the kind in question was originally picked out by ostension, as “*that kind of thing*”.<sup>8</sup> Consider the example of cats: it takes empirical investigation to learn about their underlying nature, or what kind of thing they are. For all we know a priori, it might turn out that cats are automata or demons, rather than animals. But if they share some deep explanatory nature, then whatever it

may actually be, they have *that* necessarily. We might call it their “essence”. Given that our local cats are animals, we would withhold the label “cat” from any (perhaps counterfactual) entity that lacked this internal nature, no matter how superficially cat-like it might seem. Hence ‘cats are animals’ is seen to be a necessary truth, albeit a posteriori. Similarly for theoretical identifications, such as that between *water* and *H<sub>2</sub>O*: whether the identification holds can only be decided by empirical inquiry; but if it does hold, then it does so of necessity. When assessing a counterfactual world, we judge that it lacks water if it lacks the stuff (H<sub>2</sub>O) that plays the water role in the actual world.<sup>9</sup>—

The upshot of all this is that the connection between apriority and necessity is not nearly so straightforward as claimed in CT. There are all sorts of claims – e.g. ‘water is something other than H<sub>2</sub>O’, ‘Cicero is not Tully’, or ‘Cats are strange demons’ – that are conceivably true and yet metaphysically impossible. Just because we can coherently imagine something being true, doesn’t guarantee that there is any possible world where it really is so. Modal rationalism looks to be in trouble.

### §1.3 *The Modal Rationalist’s Response*

**Modal rationalism**, recall, is the thesis that we have a priori access to the space of possible worlds. The Kripkean counterexamples to CT suggest that there are necessary truths beyond our a priori reach – statements may be true in all possible worlds without our realizing it.<sup>10</sup>— But there are two ways to account for this: perhaps we are ignorant of what modal space contains, or

perhaps we are merely ignorant of how to describe it. Only the former poses any real threat to modal rationalism; but it is the latter account that we should accept.

It is sometimes supposed that Kripke shrank metaphysical modal space so as to exclude some “conceptually possible” worlds. On this story, we imagine a conceptually coherent scenario and then realize (a posteriori) that it could not come to be. But this is not how we proceeded in the above sections at all. Instead, we imagined a scenario involving (say) cat-like demons, and then we determined that to call those creatures ‘cats’ would *misdescribe* the scenario. No total possibilities – complete scenarios or “worlds” – have been newly ruled out. Rather, certain descriptions of them have been disallowed.<sup>11</sup> The significance of the Kripkean necessary a posteriori is thus more semantic than metaphysical.<sup>12</sup> It suggests that there are true descriptions of modal space that we are not in a position to recognize a priori. But this need not involve any ignorance of modal space itself. Our ignorance may be purely linguistic.

Let me now sketch the sort of picture proposed by the modal rationalist. Modal space is taken to be rationally transparent, by which I mean that the contents of possible worlds are a priori knowable, at least in principle. An ideally rational agent could, by reason alone, come very close to omniscience. For, in a sense, the only thing the ideal agent doesn’t know a priori is *which world is actual*. Her only fundamental lack is this self-locating knowledge; from that one additional fact, she could know all. (Granted, if we reject Lewisian modal realism, we will think that the significance of actuality goes beyond mere self-location. There’s an important sense in which the actual world is the only one that’s *real*, and so this one item of ignorance

really signifies near-total ignorance of reality. Nevertheless, Lewis' picture offers a convenient heuristic for thinking about modal knowledge.)

Any truths that are not a priori knowable must, therefore, depend on this crucial unknown fact of which world is actual. Indeed, this is precisely what we find in the Kripkean cases described above. Whether a counterfactual critter counts as a cat depends on whether it shares the same underlying nature as the cat-like critters *of our actual acquaintance*. The ideally rational agent might know a priori all about the intrinsic natures of all the various cat-like beings across the possible worlds. But, due simply to her ignorance of which world is actual, she cannot know which of those beings are cats. As Jackson explains, "The key point is that the right way to describe a counterfactual world sometimes depends in part on how the actual world is, and not solely on how the counterfactual world is in itself."<sup>13</sup> The Kripkean cases arguably result from an actuality-relative semantics, whereby what (some of) our words mean depends on which world is actual.<sup>14</sup> The next section will spell out the details behind this, and thereby neutralize the Kripkean threat to modal rationalism.

#### §1.4 *Two Dimensional Semantics*

Two-Dimensionalists propose that there are two ways to think about other possible worlds.<sup>15</sup> We can hold the actual world fixed and consider another world *as counterfactual*, yielding the familiar notion of 'metaphysical', or as I will call it here, *subjunctive* possibility. Alternatively, we may consider a world *as actual*, which leads to a notion that we may call *indicative*

possibility. This duality complicates the platitude that  $S$  is necessary iff  $S$  is true in all possible worlds, for now we have two ways of assessing the truth of  $S$  at each world  $w$ . We can say that  $w$  **satisfies**  $S$  iff the subjunctive conditional “If  $w$  had been the case,  $S$  would have been the case” is true. On the other hand,  $w$  **verifies**  $S$  according to the indicative conditional: “If  $w$  is the case,  $S$  is the case.”<sup>16</sup> That is, the truth of  $S$  follows from the hypothesis that  $w$  is actual. Note that  $S$  is subjunctively necessary just in case it is satisfied by all worlds, and indicatively necessary just in case it is verified by all worlds. But these two types of necessity can diverge, as the Kripkean examples show.

Consider a possible world  $w$  that is much like ours but that the cat-like creatures in  $w$  are secretly demons. If we presuppose the actual animality of cats, and consider  $w$  as counterfactual in light of this assumption, we should conclude that there are no cats in  $w$ . However, if we consider  $w$  as actual, the situation is rather different. Upon seriously entertaining the hypothesis that  $w$  obtains, we should rationally be led to the conclusion that cats are demons. What’s more, if we consider our world  $@$  as counterfactual, under the presupposition that  $w$  is actual, then we should deny that the cat-like animals in  $@$  are real cats after all – for real cats, we suppose, are demons. The upshot of all this is that the essence of cathood is left an open indicative possibility, but once fixed, it holds of subjunctive necessity. (This is because the essential animality of  $@$ -cats, along with the essentially demonic nature of  $w$ -cats, is fixed in both senses. What’s left open is simply which of these is found in our actual cats, i.e. whether our world is  $@$  or  $w$ .)

We may generalize and precisify this account with the apparatus of two-dimensional possible-worlds semantics. We begin by associating expressions with an *extension*, or what they pick out in the world. So here ‘GWB’ and ‘The U.S. President’ both extend to Bush. But recall our earlier discussion (§1.1) of how these terms come apart in other possible worlds. This is reflected in their respective *intensions*, which are functions from possible worlds to extensions. Intuitively, we may think of intensions as determining what an expression picks out in each possible world. A rigid designator has an invariant intension. Hence, the intension of ‘GWB’ returns Bush for every possible world where he exists, but the intension of ‘The U.S. President’ is more variable, sometimes returning Kerry, for instance.

Two-dimensionalism comes in when we remember that there are two ways to consider a world. While philosophers standardly employ subjunctive or “secondary” intensions, as introduced in the previous paragraph, we might also construct a new semantic value of “primary intensions”, to mirror indicative possibility.<sup>17</sup> Here the function from worlds to extensions reflects our judgment of what the given expression picks out on the hypothesis that the world in question is actual. It is related to ‘verification’ rather than ‘satisfaction’ (as defined above). So, for instance, the primary intension of ‘cat’ picks out the cat-like creatures in each possible world. This is so even though the secondary intension of ‘cat’ is restricted to the kinds of creatures that actual-world cats are.

We may now apply this apparatus to the Kripkean necessary a posteriori: there we have claims that are subjunctively, but not indicatively, necessary. A Kripkean sentence  $S$  is satisfied at all worlds although its negation,  $\sim S$ , is

verified at some worlds. *S* thus has a necessary secondary intension, but a contingent primary intension. All this derives from the simple fact that *S* is true at all worlds considered counterfactually, but *not* at all worlds considered as actual. This explains why it is subjunctively (or “metaphysically”) necessary without being a priori. There are ways the world could be such that, if the world *is* that way, then *S* is false. It takes empirical investigation to be sure that the antecedent here fails. But given that this is actually the case, the truth of *S* then holds of subjunctive necessity.

Recall the intuitive argument for CT presented in the introduction. There it was suggested that a claim should be a priori just in case it will be true no matter how the world turns out, i.e. necessarily true. This ran into problems because our actuality-relative semantics entails that whether a claim is true of a counterfactual world may depend on the a posteriori fact of which world is actual. But we are now in a position to see that the original argument can be saved, it simply needs to be interpreted in the indicative mood rather than the subjunctive. There may be a posteriori claims true of all worlds considered counterfactually; but apriority should instead be linked to truth in all worlds considered as actual – that is, verification by all worlds, rather than satisfaction by all. We may thus revise the apriority-necessity coincidence thesis as follows:

**(CTI):** For any statement *S*: *S* is a priori iff *S* is indicatively necessary.

CTI is still far from trivial, and will be explicitly challenged in later chapters. But for now it suffices to note that the standard Kripkean cases offer no counterexample to it. Further, CTI paves the way for modal rationalism.

Indicative possibility reflects the space of possible worlds, albeit under a certain “mode of presentation”, i.e. their being considered as actual. So a priori access to indicative modal truths entails a kind of a priori access to modal space itself. The nature of this connection can be further clarified by the notion of semantic neutrality. Following Chalmers (2002), let us say that an expression is *semantically neutral* just in case its primary and secondary intensions coincide. Intuitively: the expression is unaffected by whether we consider worlds as actual or as counterfactual. Hence, a neutral statement will be indicatively necessary just in case it is subjunctively necessary. So even the original CT will at least hold true of this restricted class of statements. (Examples include descriptive terms such as ‘cat-like creature’ or ‘watery stuff’, as opposed to the rigidified ‘cat’ or ‘water’.)<sup>18</sup>— We may thus restate it:

**(CT\*):** For any *semantically neutral* statement  $S$ :  $S$  is a priori iff  $S$  is necessary.

Finally, it is plausible that any possibility can be (re-)described in semantically neutral vocabulary, thereby providing us a priori access to it. Modal rationalism is thus restored.

## Chapter Two

### §2.1 *The Idea of “Strong Necessities”*

We have seen that standard examples of the necessary a posteriori pose no threat to modal rationalism. They don't involve any shrinking of modal space, and can be explained away as involving semantic rather than metaphysical ignorance. For all that has been said so far, it remains plausible that each coherent claim (i.e. that cannot be ruled out a priori) has a non-empty primary intension, or is verified by some possible world, in the sense that the claim is true of that world considered as actual. This may be understood as the central claim of modal rationalism: for every conceptually coherent scenario, there is a possible world to match.<sup>19</sup>—

This is the claim that must be denied by the opponent of modal rationalism. They must hold that there are not enough possible worlds to go around. This is what's needed to break the link between apriority and indicative necessity. A claim could then be true of all possible worlds – whether considered as actual or counterfactual – without being a priori, because there is a coherent yet strictly impossible scenario that purports to falsify the claim. We may call a claim with these modal properties a *strong necessity*.<sup>20</sup>— These amount to a posteriori necessities with a necessary primary intension: not only are they true in all worlds considered counterfactually, but also in all worlds considered as actual. Despite this, they fail to be a priori because we cannot know a priori that the worlds it is true of (considered as actual) really exhaust the possible worlds. We may think that there are others – we can imagine

coherent scenarios that falsify the strong necessity – but those scenarios we imagine fail, without our realizing it, to correspond to any genuinely possible world.<sup>21</sup> – Not all rationally apparent possibilities are real possibilities.

An example of this position could be found in a theistic view according to which God is a necessary being even though his non-existence is conceptually coherent and so cannot be ruled out a priori.<sup>22</sup> – We can imagine all sorts of scenarios that don't contain any deities, but this theist will deny that they correspond to any genuinely possible worlds. They will insist that God exists in all possible worlds, and our failure to grasp this is due to the rational inaccessibility of modal space. A priori reflection does not suffice to establish what possibilities there are. Hence we might mistakenly conceive of Godless worlds without appreciating the brute fact of their intrinsic impossibility.

This anti-rationalist thesis can be further clarified in terms of Chalmers' construction of an "epistemic space" of a priori coherent scenarios. Without getting into the details here, we may consider these scenarios to be equivalence classes of maximally consistent sentences in an idealized language, individuated by their a priori implications. We have a priori access to this epistemic space, since it is explicitly defined in terms of rationally coherent possibilities: for every claim that cannot be ruled out a priori, there is a scenario which verifies this claim. Further, the idealized language is – modulo 'centering' indexicals – semantically neutral, so that we do not need to know which world is actual in order to fully grasp the meanings of the terms or know how to apply them to other situations. This entails that truth is *scrutable* in epistemic space, in that we have a full a priori grasp of what is

true in each scenario. This yields a sort of conditional knowledge: for any statement S that's true in a scenario canonically described by D, 'D implies S' is a priori knowable.

So we have a priori access to epistemic space in two respects: firstly, assuming that the range of a priori claims is itself a priori knowable, we can know what scenarios there are; and second, by the scrutability thesis, we can know what is true in each of them. For a claim to be true in all scenarios is just for it to be a priori, or *conceptually necessary*. If it is true in some scenario, then it is conceptually possible, and cannot be ruled out a priori. These latter claims follow from the definition of scenarios in terms of coherent possibilities. In effect, a scenario *just is* a complete claim about the world that cannot be ruled out a priori. So if S is true in all scenarios, it is part of every complete claim that cannot be ruled out a priori, and so its falsehood must be something that *can* be ruled out a priori. Hence the identification of conceptual necessity with truth in all scenarios.

We are now in a better position to clarify the dispute between modal rationalists and the defenders of strong necessities. Modal rationalists hold that epistemic space maps onto metaphysical modal space.<sup>23</sup> Our a priori grasp of the former is thus transmitted to the latter. If scenarios *just are* (centered) possible worlds, then we can likewise know what worlds there are, and what semantically neutral claims are true in each of them. The anti-rationalist denies all this. If we grant the construction of scenarios, the anti-rationalist must insist on a separation between these and possible worlds, so that the latter remain beyond rational reach. This then makes it clear how strong necessities can arise: though true in all worlds, there is a falsifying

scenario whose possibility we cannot rule out a priori. The space of genuinely possible worlds is narrower than the space of conceptually possible scenarios – and it is only the latter to which we have rational access.

Our central dispute thus concerns how many spaces we need in order to ground the epistemic and metaphysical modalities. The modal rationalist proposes that a single space of worlds/scenarios will do the trick: as outlined in Chapter One, “conceptual” possibility and “metaphysical” possibility correspond, respectively, to the indicative and subjunctive modes of assessing a single, unified space of worlds. The anti-rationalist denies this by suggesting that there are not enough worlds to go around. In effect, they require an *independent* space of conceptually possible scenarios, broader than the space of ways the world really could have been. Put another way, since both parties at least agree on the scope of conceptual possibility, perhaps the disagreement is better characterized as whether we need an independent – and presumably more restricted – space of *metaphysically* possible worlds. The modal rationalist proposes that our epistemic space suffices to ground metaphysical modality, whereas the anti-rationalist insists that a distinct space of worlds is required here.

Modal rationalism offers the more parsimonious account, and so is preferable so long as it proves theoretically adequate in all other respects. Indeed, the rationalist’s main argument against strong necessities is that they’re unmotivated: we simply don’t have any reason to believe in such things.<sup>24</sup>— They involve the positing of extra theoretical complexity, in the form of a further primitive modal space, and we should be reluctant to go to such lengths without good reason. Chapter Three will explore this notion of

“primitive modality” in more detail. But first, I will discuss anti-rationalist arguments that purport to undermine modal rationalism from within.

## §2.2 *The Challenge from Meta-modal Conceivability*

Modal rationalism could be shown to be internally inconsistent if there were individually coherently conceivable claims that gave rise to incompatible commitments concerning the entirety of modal space itself. Consider, for example, the thesis that a necessary being – call it ‘God’ – exists.<sup>25</sup> Yablo (1999) proposes that this thesis and its negation are each coherently conceivable.<sup>26</sup> But it follows from S5 modal logic that they cannot both be possible. Let G stand for the sentence ‘God exists’, which – given the definition of God as a non-contingent being – is interchangeable with  $\Box G$ . I will use ‘ $\blacklozenge$ ’ as the ideal conceivability operator, so that ‘ $\blacklozenge P$ ’ means that P is coherently conceivable when stated in semantically neutral terms. Then we have the standard operators: ‘ $\diamond$ ’ denoting metaphysical possibility, ‘ $\Box$ ’ for metaphysical necessity, ‘ $\sim$ ’ as negation, ‘ $\rightarrow$ ’ for material implication, and ‘ $\&$ ’ for conjunction. Then we have the following quick proof:

1.  $\blacklozenge \Box G \ \& \ \blacklozenge \Box \sim G$  (conceivability premise)<sup>27</sup>
2.  $\blacklozenge P \rightarrow \diamond P$  (conceivability-possibility axiom, for reductio)
3.  $\blacklozenge \Box G \rightarrow \diamond \Box G$  (substitution, 2)
4.  $\diamond \Box G \rightarrow \Box G$  (S5 axiom)
5.  $\Box G \rightarrow G$  (T axiom)
6.  $\blacklozenge \Box G \rightarrow G$  (3, 4, 5, transitivity)

7.  $\blacklozenge \Box \sim G \rightarrow \sim G$  (derive similarly to the above)
8.  $G \ \& \ \sim G$  (from 1, 6, 7, *modus ponens*) – **contradiction**.

To explain this in intuitive terms: the only way it can *anywhere* (i.e. in some possible world) be true that  $G$  is true *everywhere* (in all possible worlds), is for  $G$  to really be true everywhere. Hence possible necessity implies actual necessity. But then, if conceivability implies possibility, then the transitivity of implication entails that conceivable necessity implies actual necessity – and hence actual truth. So if each of two incompatible necessity claims are individually conceivable, then modal rationalism will straightforwardly lead us to a contradiction.

The point can be made in an even more general way, for what S5 captures is the idea that modal space as a whole is static, fixed and necessary.<sup>28</sup> Contingent truths may vary from world to world, but the entirety of modal space remains unchanged no matter your vantage point. (Compare: dynamic truths may vary from time to time, but the timeline as a whole is eternal.) Modal space is what it is necessarily: though our world could have been different, the total sum of possibilities could not. So if there are multiple conceivable ways for modal space to be, no more than one of them can be genuinely possible. The others will serve as counterexamples to modal rationalism, in virtue of being coherently conceivable and yet metaphysically impossible.

In the case discussed above, it is supposed that both a modal space where God exists in every possible world, and a modal space in which he does not,

are each coherently conceivable. But the challenge could be put even more directly. It might be suggested that modal rationalism as a thesis is itself conceivably false.<sup>29</sup>— Though modal rationalism claims that metaphysical modal space is as broad as epistemic space, it might be thought that we can easily imagine the contrary. It seems conceivable that epistemic space might outstrip metaphysical modal space, so that there are coherently conceivable semantically neutral claims that are nonetheless metaphysically impossible. But if this meta-modal claim is possibly true, then it must – by S5, as explained above – be actually true, hence refuting modal rationalism. What response can the modal rationalist make here?

One option would be to reject S5 – or premise (4) in the formal argument above – and instead hold that modal space as a whole really could have been different. This is a large bullet to bite, and arguably would not help the modal rationalist in any case. If there are multiple possible modal spaces, depending on which world is actual, then we presumably lose our a priori grasp of modal space. In order to discern what's actually possible, we would first have to determine which world is actual – a job for empirical science. At best, we might develop a “meta-modal rationalism”, granting us an a priori grasp of the various possible modal spaces, without knowing which describes the actual possibilities. But if this thesis is also conceivably false, we risk infinite regress.

Alternatively, one might preempt the substitution in (3) by further restricting the applicability of the conceivability-possibility axiom.<sup>30</sup>— This would lead to a *weak modal rationalism* that proposes a link between apriority and necessity only for first-order, non-modal claims. By denying that meta-modal

conceivability implies possibility, this revised view clearly escapes Yablo's arguments. The revision is not entirely *ad hoc* either, since the modal nature of meta-modal claims is clearly very different in kind from standard first-order claims. This is, after all, precisely why distinctive problems arose. So there's a natural distinction to be drawn here, and the resulting thesis is still interesting enough in its own right, even if it is less strong than might be hoped for.

Nevertheless, I agree with Chalmers (1999) that the most attractive response for the modal rationalist here is to hold on to their strong position, and instead deny the meta-modal conceivability intuitions found, for example, in premise (1) above. It isn't at all clear that a necessary being, or a shrunken modal space, *is* coherently conceivable in the appropriate sense. The modal rationalist will want to hold that their position is not just true, but *a priori*. They would then expect opposing views to be refutable a priori, and hence not feature in any a priori coherent scenario. Of course, it would beg the question to merely assert: "the thesis is true and hence has no successful counterexamples". But that is not what's going on here. Rather, I hope to show that the modal rationalist can explicate their commitments in a way which makes clear exactly why, on their view, the meta-modal cases in question are not taken to be genuinely conceivable. If successful, this should suffice to undermine the charge of internal inconsistency or self-refutation.

The appropriate sense of *conceivability* can be found in Yablo's own (1993). He suggests that, rather than attempting to assess a claim in abstract isolation, one must imagine a scenario that verifies the claim. It is whole individual scenarios that are fundamental for conceivability claims. We cannot directly

conceive, in the relevant sense, of an isolated claim. Rather, we must conceive of a whole scenario that we take to verify the claim in question. In this respect, Yablo works down from scenarios to the first-order claims that they verify. But we might also work upwards, from scenarios to the meta-modal claims that are true of all of them. This is likewise justified by the principle of taking scenarios as fundamental. I propose that we cannot *directly* conceive, in the relevant sense, of an entire modal space. Rather, we must conceive of the individual scenarios that we take to comprise the space of possibilities.

We may formalize this claim as the following cross-modal variation on the S5 axiom [where  $\blacksquare P = \sim \blacklozenge \sim P$ ]:

$$(S5^*): \blacklozenge \square G \rightarrow \blacksquare G$$

On this proposal, for a necessary being to be coherently conceivable, it must be the case that the being exists in every scenario that can be conceived. The failure of the latter condition suffices to render the necessary being inconceivable. (Assume I can coherently conceive of a Godless world. That possible world doesn't disappear when I turn to the question of whether God might necessarily exist. It can legitimately influence my modal theorizing. In particular, it precludes the possibility of God existing in *every* possible world, for I can see all along that he doesn't exist in *that* one, at least.) More generally: there are not multiple conceivable modal spaces, for that would require, impossibly, that there be more than one totality of individually conceivable scenarios.

The anti-rationalist might complain that my response here is still question-begging. After all, they contend that there are coherent scenarios which posit different respective modal spaces. By proposing that meta-modal conceivability is a unique construction from the space of individually conceivable scenarios, I have effectively ruled out their contention from the start. Ideally, we should want some independently compelling explanation of why meta-modal conceivability should be constructed bottom-up from individual scenarios in the way I have suggested. A conclusive answer must wait upon a proof of modal rationalism itself, for what we need is to show that the alternative meta-modal claims can be ruled out a priori, and hence feature in no coherent scenario. Such a project goes well beyond the scope of this paper. But I hope to have at least blunted the force of the anti-rationalist's challenge, by indicating a plausible line of response that is open to the modal rationalist.

The challenge from meta-modal conceivability rests on assumptions that the modal rationalist need not grant, and so is question-begging in its own right. We may be left in a state where, due to the absence of shared premises, neither party can rationally convince the other to change their position. Even so, at least this counteracts the claim that modal rationalism is *internally* inconsistent.

### §2.3 *The Challenge from Unknowable Necessities*

The claims of such paradigmatically *a priori* disciplines as mathematics and metaphysics are typically thought to be non-contingent, and hence necessarily true if they are true at all. But it isn't clear that all such truths are *a priori* knowable at all. Perhaps, even after idealizing away our contingent cognitive limitations, there would still be no way to establish either Goldbach's Conjecture or its negation, even though one or other of them must be true in fact, and hence necessary.<sup>31</sup>— This would then provide a counterexample to modal rationalism, by way of a semantically neutral necessary truth that nonetheless fails to be *a priori* knowable. While the modal rationalist takes necessities to be truths of reason, the present challenge proposes that there may be necessary truths beyond the reach of reason.

The challenge is limited, because there are no clear cases of such unknowable truths, nor any good reason to believe that any such exist. Here one might mention Gödel's incompleteness theorems, but despite the popular misconception, Gödel merely establishes limitations on what can be proved within particular formal systems, rather than proving that there are mathematical truths that we cannot come to know by *any* rational means. And any particular trouble cases – e.g. Goldbach's Conjecture – might well be soluble in principle, for all we presently know. So there is no knockdown objection to modal rationalism being offered here. What the challenge does achieve, I think, is to remind us that we can make sense of the idea of a gap between necessity and *a priori* knowability. They don't appear to be the exact

same thing. This idea motivates the modal primitivism, to be discussed in the next chapter, that is a precondition for strong necessities.

## Chapter Three

### §3.1 *Metaphysical Realism and Conceptualism*

Modal rationalism links metaphysical necessity to a priori knowability. We may wonder what this implies about the metaphysical status of modal discourse: can it still be fully mind-independent? The modal rationalist grants that many modal facts will never be actually known – and perhaps even *cannot* be known by creatures with our cognitive limitations. So the modal facts are genuinely objective, in that they are completely independent of *our* minds, and may transcend at least the evidence that is practically available to us. Nevertheless, modal rationalists hold that the sum of all *possible* rational evidence, including that which is accessible only to more cognitively advanced agents, suffices to settle the modal facts. At the end of the day, all (semantically neutral) necessary truths must be knowable on ideal rational reflection. There are no such necessary truths besides those that are so knowable. Modal reality cannot transcend all possible rational evidence.

Why not? Here modal rationalists may split into two camps. *Conceptual* modal rationalists, e.g. Chalmers (1999), seek to epistemicize modality by claiming that so-called “metaphysical necessity” is really nothing over and above a priori knowability.<sup>32</sup>— On this view, there is an analytic link – perhaps identity – between the concepts of possibility and ideal conceivability that precludes any gap between the two. That way, modal truth *just is* the ideal limit of a priori inquiry; it does not answer to the sort of independent reality that might sensibly be considered beyond all epistemic reach.<sup>33</sup>— Modal reality

is thus a kind of (non-contingent) rational construction. Rather than addressing the metaphysical question of how reality is in itself, modal facts may be considered more fundamentally normative in nature: they tell us what should be concluded at the ideal limit of a priori rational reflection.

*Metaphysical Realists* about modality, in contrast, wish to uphold the conceptual distinction between necessity and apriority, whereby the former is taken to be a genuinely metaphysical notion – about the world as it is in itself, rather than our (even idealized) beliefs about it. Realists will be more sympathetic to the previous chapter's conclusion (§2.3) that we can make sense of a 'gap' between the two concepts, according to which even ideally rational agents might be inescapably mistaken about the breadth of modal space. Realist modal rationalists must simply insist that there is no such gap as a matter of fact. Further, this fact about modal space will be necessary if true at all (cf. §2.2), so the modal rationalist is also committed to its apriority. But whereas the conceptualist takes the connection to be analytic, realists will instead propose that it is a substantive, synthetic claim about the metaphysical nature of reality. It is this 'realistic' version of modal rationalism that I will seek to elucidate and defend later in this chapter. First, we must depart from the conceptualists by taking seriously the idea of primitive metaphysical modality that underlies the radical challenge from strong necessities.

Note that metaphysical realism puts modal rationalism at risk by opening the door to strong necessities. If – contrary to my arguments – there are indeed some strong necessities, then the inference from ideal conceivability to metaphysical possibility is jeopardized. Even so, this might not leave

philosophers quite as hamstrung as typically supposed. It is arguably the fundamentally *rational* notion, rather than the *metaphysical* one, that we employ in our philosophical theorizing. That is, for the standard theoretical uses of modality, it may be the conceptualist's space of coherent scenarios that we really need.<sup>34</sup>— But I think we have a grip on an independent metaphysical notion in any case, so I will try to bring this out in the sections that follow. My subsequent defense of realist modal rationalism will be of greater significance to those who dispute the theoretical primacy of epistemic space proposed above.

### §3.2 *Content-Based Modalities vs. Metaphysical Modality*

I wish to distinguish two very different ways of specifying a modal space. In the first case, philosophers may isolate and identify the particular modal space they wish to work with by offering a (more or less) formal specification of the contents they wish to include or exclude. That is, they begin with some framework *F* of rules or limitations, and then define the space of *F*-possibilities as simply a matter of what is *not ruled out* by *F*.<sup>35</sup>— I will say that such a specification is “content-based”, to highlight the point that its delimiting rules are directly and exclusively concerned with the internal contents of possible worlds, so that one may determine whether or not to allow a world-candidate solely on the basis of descriptions of *what that world contains*. For example, nomological possibility is sometimes understood as simply consisting in the non-violation of the actual laws of nature.<sup>36</sup>— This specification is content-based insofar as it can be applied simply by examining a complete description of the internal workings of a candidate world, and determining whether any of the described events contravene our

laws of nature. A claim is nomologically “possible”, in this sense, so long as the laws of nature don’t rule it out. Conceptual possibility can similarly be settled simply by determining whether a candidate world-description contains any overt or implicit self-contradictions. This might not be purely formal: if rational insight cannot be captured algorithmically, there will be no finite set of rules that can determine a priori coherence. So the latter may need to be taken as a primitive in its own right. Nevertheless, this modality is “content-based” as I use the term, for it serves to directly fix the breadth of the modal space, by means of a particular method.

In this chapter, I wish to explore the proposal that metaphysical possibility is *not* to be understood in such content-based terms. Its breadth is not defined in terms of epistemic procedures. Instead, this modal space is, in a sense, “world-oriented”. It is to be characterized first and foremost in terms of its *metaphysical* nature, thus leaving its breadth of content to be fixed by reality rather than built explicitly into the concept. This aims to connect with our intuitive notion of ‘metaphysical possibility’ as reflecting ways the world *really* could have been – a concept whereof our primitive grasp leaves open, at least initially, what breadth of content this modal space contains. The answer is fixed by reality, not our concepts alone. The question of *what really could have been* is here assumed to be a question fundamentally about the world – or reality in itself – that admits of an objective and exclusive answer. Though philosophers might propose whatever content-based restrictions suit their purposes, the world itself provides just one space of *real* possibilities.

Of course, this space of metaphysically possible worlds must have some or other breadth, and so be specifiable in terms of restrictions on content.

Perhaps it includes all the conceptually possible worlds. Or perhaps it includes only the actual world, as would be the case if things never really had any genuine opportunity to be different. If all else fails, one could simply give an exhaustive specification of each and every world that it includes, and construct their disjunction as the content restriction. Any of these “spatial breadths” are *prima facie* consistent with the concept I’m trying to point to, because its fundamental character lies in a different dimension. Unlike the other modal concepts, we don’t immediately characterize it in terms of breadth or restrictions on content. The criterion for a world’s inclusion in this space is instead its brute modal nature. We don’t ask: “Does this world contain anything which violates such-and-such content restrictions?” Instead we ask the irreducibly modal question: “Is this a world that had the opportunity to be actualized?” Or, equivalently: “*could it really have come about?*”

Once we have grasped metaphysical modal space by way of the above questions, we can go on to inquire into the space’s breadth of content – as below. But for now I emphasize that the concept must be initially grasped in these primitive modal terms. You cannot *begin* by characterizing metaphysical modal space in terms of its contents, because those are not included in the concept as it initially presents itself to us. If you begin with them, you are really grasping a different concept altogether. After all, for any space of worlds characterized in terms of their content, one can still coherently ask: “but might they really have been actualized?” It wrongly remains an open question, unless one builds this modal requirement right into the fundamental conceptual character of metaphysical modality.

Note that this conception makes no explicit demands on what content must be found within candidate worlds. Metaphysical modality, thus understood, is not to be analyzed in terms of any collection of formal rules or laws that must be satisfied, nor even a primitive content restriction such as “rational coherence”. Rather, what matters is simply whether the candidate world is one that *really could* have been actualized. So long as this external modal property is satisfied, we need not worry about what is *in* the candidate world. This makes the specification of metaphysical modal space significantly different in kind from the content-based spaces mentioned earlier.

### §3.3 *Identifying Metaphysical Modality*

It might be wondered what, exactly, this notion of “really could have been actualized” involves. (Merely emphasizing the “*really*” will do little to help one who lacks an antecedent grasp of the concept.) Since it is presented as a primitive or bedrock concept, no reductive analysis can be offered to explicate it. But some general remarks may help bring the intuitive notion to light.

Chalmers expresses his skepticism as follows:

It seems to me that we do not even have a distinct *concept* of metaphysical necessity to which the second primitive [besides rational coherence] can answer. The momentary impression of such a concept may be a residue from initial impressions of the Kripkean distinction between epistemic and metaphysical modality. But once we recognize that this distinction can be explained with one modal primitive, and

that there are constitutive ties between the Kripkean modalities, the grounds for this impression disappears. The only concept of a "metaphysical possible world" that we have is that of a logically possible world. If someone thinks they have a distinct concept here, there is no reason to believe that anything answers to it.—<sup>37</sup>

I think both challenges may be answered by pointing to a familiar – and metaphysically ‘realist’ – concept of possibility that is sometimes relegated to the merely “nomological”. The concept I have in mind naturally relates to commonplace ideas about objective chance, indeterminism, and the open future. Many people think that the future is metaphysically open, in that it *really could* turn out in any one of a number of different ways – the truth of the matter hasn’t been decided yet. Each open possibility has some non-zero objective chance of eventuating. Note that this isn’t just a claim about our epistemic situation, or even the rational ideal: it’s about how reality is in itself. This has nothing to do with any actual or possible minds. As a rough heuristic: if God were to rewind time and play it back again, things would unfold differently. Admittedly, this commonsense belief assumes indeterminism. If, instead, the future is already determined, then – given the present state of affairs – there is only one way that things can really turn out. No matter how many times God “replays” history from this point, he’ll never get a different result.

Of course, metaphysical possibility cannot simply be identified with non-zero objective chance. The past is presumably now fixed, so there’s no chance it will suddenly change on us. But even though the past certainly *won’t* be different, nevertheless we might still think that it *could have* differed.—<sup>38</sup>

Perhaps there were open alternatives at a time in the even more distant past. In extending our intuitive notion of the open future back into the past, we will find various (now closed) branches that really were, at one point, dynamically open possibilities. Our concept of metaphysical possibility should at least include the entire history of such dynamic possibilities.<sup>39</sup>— They are all ways that the world really could have turned out. Hopefully it is now intuitively clear what I mean by this.

It's worth pausing here for a moment to clarify what we have established. In exploring the metaphysical specification of modality, we have thus far reached a space of worlds that could be given the content-based specification of “nomological possibilities given the initial starting conditions of the universe”. But such content-based descriptions fail to capture the metaphysical significance – the idea that the world *really could* have turned out in any of those ways. (Though readers might implicitly project significance on to it, in light of their background knowledge that anything satisfying this content restriction would in fact have had an objective chance of eventuating.) I wish to draw attention to this modal primitive – the one we invoke when thinking about objective chances, physical indeterminism, and open futures – and how natural it is for us to be Metaphysical Realists about it. This seems to be a species of modality that is no mere rational idealization, but rather is truly *in the world*, as a basic component of reality.

It might be objected that what I've pointed to here *just is* “nomological modality” in some sense. Non-Humeans, at least, could be expected to imbue some form of natural necessity with the primitive metaphysical significance proposed above – in which case the end result will be much the same.<sup>40</sup>—

Whatever you want to call it, once we have a grip on this ‘realist’ modal primitive, we may ask: could the laws of nature themselves have been different, in this primitive sense? It seems a reasonable question, though of course the nomological impossibility of it is trivial. This suggests that the primitive notion in this vicinity is not, strictly speaking, nomological possibility after all – at least, it doesn’t seem built into the modal concept that the laws of nature must be necessary in this primitive sense.<sup>41</sup>— We can always stipulate such a content-based restriction later, if we need it for other purposes. But the core concept here is – *prima facie* – potentially broader than that. So I think ‘metaphysical possibility’ is the more fitting term.

We have here a metaphysically ‘realist’ modal concept that has worldly application: at the very least, it spans the entire history of dynamic possibilities or “open futures”. This undermines the skeptical basis for conceptualism – we should be realists about metaphysical possibility instead. But how far does it extend? We are now faced with the awesome questions of why our universe exists at all, and whether a wholly different universe – say with alternative laws of nature or initial conditions – could have existed in its place. This provides the focus for the next section. Could absolutely any coherent scenario really have been actualized, as the modal rationalist proposes? Or are some rationally apparent possibilities necessarily excluded by the nature of reality, creating “strong necessities” that fall outside the 2-D framework and falsify modal rationalism?

### §3.4 *Two Principles of Modal Expansion*

To recap: we have a space of scenarios, each of which represents a way for the world to be. We might think of each of those “ways” as being a maximal property, just one of which is instantiated by the actual world, and hence is “the way the world is”.<sup>42</sup> That these various properties exist is, it seems, a merely ontological fact.<sup>43</sup> What we’re interested in is the *modal* fact concerning which of these rival properties (scenarios) had a real chance to be instantiated (actualized). The answer will yield our space of metaphysically possible worlds. Any leftover scenarios will be coherent ways for a world *to* be, but nevertheless the world *could not* really have been such a way. They are the ways that “never stood a chance”, so to speak. It might already be thought that there’s something very strange about the idea of such leftovers. Let me offer further grounds for such skepticism.

The above discussion frames the modal question in terms of a *positive* demand for some reason to think that a scenario had the opportunity to be actualized. This presupposes that scenarios are “modally inert” by default. Their being is merely ontological, and some further modal property or relation needs to be *added* to them in order to make them “really possible”. They must be targeted by some potentially world-actualizing mechanism – perhaps a Leibnizian God who surveys the space of possible worlds before deciding which to bring into actual existence. But this notion of an atemporal process of worldly “becoming” is of dubious coherence. There is no time before time began, during which such a process of selection could take place. We might take this to indicate the metaphysical necessity of our actual laws and initial conditions – thus contradicting modal rationalism. Or we might reframe the question in a way that escapes these problems. Here I seek to explore the latter option.

We might achieve this by framing the question in *negative* terms, or asking whether there is any reason why a described world-candidate could *not* have been actualized. Here we treat possibility as the default assumption: absent any reasons to the contrary, we assume that each way for a world *to* be is indeed a way that the world really *could* have been. So long as there is nothing necessarily preventing a candidate from being actualized, it should thereby be considered possible. It does not require any positive mechanism that could have brought it about, or “given it a chance”. No such chance need be *given*; rather, it comes for free. We might say that a candidate’s natural state is possibility – additional reasons are required to *preclude* its possibility, not to *grant* it. Let us call this principle **the presumption of possibility**: any world-candidate will be metaphysically possible, unless there is an explicit reason why it could not have been brought about. This principle may allow us to pursue modal inquiry whilst avoiding the confusion inherent in positive demands for a world-creating mechanism.

The negatively framed question also seems more susceptible to being answered. The question of what might bestow metaphysical possibility on a world-candidate seems hopelessly mysterious. But if we ask what kind of thing could preclude a claim from being possible, we find an obvious answer, namely, *incoherence*. **The coherence principle** claims that this is the only answer available, so that a priori incoherence is the only barrier to possibility. After all, we have no trouble granting that a logically inconsistent state of affairs could not have obtained. But why think that a coherent scenario could not really have been actualized? Such a proposal would seem entirely unmotivated – there is nothing intrinsically disqualifying about the scenario,

and there doesn't appear to be anything external necessarily preventing it from being actualized either. So it seems most reasonable to conclude that any coherent scenario really could have been actualized after all. The combined effect of my two "expansionist" principles – the presumption of possibility, and the coherence principle – is to lead us back to modal rationalism, only this time with a realist metaphysical foundation.

Although the proposed principles seem quite plausible to me, they could reasonably be denied. Of particular concern is the idea that nothing is "necessarily preventing" coherent scenarios from being actualized. Given that a scenario is in fact not actualized, we might wonder why that is. Whatever the actual reason why some other conceptually possible universe does not exist in place of ours, perhaps this very same reason holds of necessity, so that the other universe *could not* really exist in place of ours. If we take actual existence to be a matter of brute fact, why not metaphysically possible existence likewise? (Perhaps they come down to one and the same fact, viz. our universe's origin.) Whereas I formulated the above principles in order to sidestep unanswerable questions about ultimate origins, critics might consider *stopping right there* to be the more appropriate response. As indicated earlier, the metaphysical realist might conclude that the origin of our universe could not really have been any different. I have shown how modal rationalists might hope to avoid this result, by formulating plausible principles of modal expansion.<sup>44</sup>— But a more thorough defense of these principles awaits further work.

## Conclusion

It's natural to expect that *what can be known without needing to look at the world* is closely tied to *how the world metaphysically could or must have been*. If we can only learn a fact a posteriori, through empirical investigation, we may expect that this is because there are other possible worlds in which the fact in question fails to hold. Assuming that possible worlds are wholly self-contained, we would not expect that examining the actual world could tell us anything informative about *other*, non-actual possibilities. Modal rationalism draws on these intuitive ideas by positing an intimate link between apriority and necessity, according to which an ideally rational agent could in principle grasp modal space – or apprehend what is possible and what is not – through the exercise of reason alone.

Kripke's discovery of the *necessary a posteriori* casts doubt on this picture. There are some necessary truths – e.g. 'water is H<sub>2</sub>O' – which can only be known after empirical investigation. But the modal rationalist suggests that the problem here is merely semantic. We can know a priori how all the various possible worlds are *in themselves*; what we don't always know is how to apply our words to them. Some terms, like 'water', are not semantically neutral – their application to counterfactual worlds is contingent on how the actual world turns out. That's why empirical inquiry may be required before we can accurately assess various modal claims. The extra work is required to grant us semantic, not metaphysical, knowledge. We may avoid this need by restating a claim in neutral terms, for which the semantic values are unaffected by whether we consider a world "as actual" or "as counterfactual".

Chapter One thus established that the Kripkean challenge to modal rationalism is toothless after all; the link between apriority and necessity may be restored by restriction to semantically neutral vocabulary.

What's needed to refute modal rationalism are "strong necessities", i.e. claims that are true in all worlds considered as actual, despite being conceivably false. This requires that there be coherent scenarios that would not be verified by any possible world. Chapter Two explored this idea further, and assessed Yablo's arguments for the claim that modal rationalists must recognize such strong necessities. Arguments from meta-modal conceivability provide the greatest challenge here, but I proposed that modal rationalists should respond by treating scenarios as epistemically fundamental, so that meta-modal conceivability is then uniquely determined by the sum of individually conceivable scenarios. Other arguments assume that there are unknowable necessities – an assumption we have no reason to grant, but that at least suggests the intuitive need for a non-epistemic foundation to modality.

Chapter Three set about exploring this idea further. I presented a metaphysically 'realist' understanding of metaphysical modality, and defended it against the conceptualist's skepticism by highlighting its connection to our intuitive ideas about physical indeterminism, objective chance, and the open future. The realist's primitive conception of modality forces them to take seriously the idea of strong necessities, but they need not give up on modal rationalism altogether. I suggest two principles of modal expansion – the presumption of possibility, and the coherence principle – which together serve to ground modal rationalism on a realist foundation.

The end result is, I think, an attractive and defensible view, which preserves many of the intuitive claims we would wish to make about modality. And although it is arguably the conceptualist's epistemic space that matters for key theoretical purposes, many would dispute this claim – which cannot be fully defended here – and so it is worth establishing the viability of realist modal rationalism for those who would place greater weight on this metaphysical modal space.

## Bibliography

Bonjour, L. (1998) *In Defense of Pure Reason*. Cambridge: Cambridge University Press.

Chalmers, D.J. (ms.) *The Multiplicity of Meaning*.

Chalmers, D.J. (forthcoming) 'Two-Dimensional Semantics' in Lepore and Smith (eds.) *Oxford Handbook to the Philosophy of Language*.

Chalmers, D.J. (2002) 'Does Conceivability Entail Possibility?' in T. Gendler & J. Hawthorne (eds.) *Conceivability and Possibility*. Oxford: Oxford University Press.

Chalmers, D.J. (1999) 'Materialism and the Metaphysics of Modality'. *Philosophy and Phenomenological Research* 59: 473-93.

Chalmers, D.J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Davies, M. & Humberstone, I.L. (1981) 'Two notions of necessity' *Philosophical Studies* 58: 1-30.

Edgington, D. (2004) 'Two Kinds of Possibility' *Supplement to the Proceedings of The Aristotelian Society* 78: 1-22.

Fine, K. (2002) 'The Varieties of Necessity' in T. Gendler & J. Hawthorne (eds.) *Conceivability and Possibility*. Oxford: Oxford University Press.

Howell, R. (ms.) 'The Two-Dimensionalist *Reductio*'.

Jackson F. (1998) *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.

Kripke, S. (1980) *Naming and Necessity*. Oxford: Basil Blackwell.

Lewis, D. (1986) *On the Plurality of Worlds*. Oxford: Basil Blackwell.

Moore, G.E. (1903) *Principia Ethica*. Cambridge: Cambridge University Press.

Putnam, H. (1975) 'The Meaning of 'Meaning'' *Mind, Language and Reality*. Cambridge: Cambridge University Press

Sidelle, A. (1989) *Necessity, Essence, and Individuation: a defense of conventionalism*. Ithaca, N.Y.: Cornell University Press.

Soames, S. (forthcoming) 'Kripke on Epistemic and Metaphysical Possibility' in A. Berger (ed.) *Saul Kripke*.

Stalnaker, R. (1976) 'Possible Worlds' *Nous* 10: 65-75.

Sturgeon, S. (forthcoming) 'Apriorism about Modality' in Hale *et al* (eds.) *Modal Content & Modal Knowledge*. Oxford: Oxford University Press.

Van Inwagen, P. (1998) 'Modal Epistemology' *Philosophical Studies* 92: 67-84.

Yablo, S. (1999) 'Concepts and Consciousness' *Philosophy and Phenomenological Research* 59: 455-63.

Yablo, S. (1993) 'Is Conceivability a Guide to Possibility?' *Philosophy and Phenomenological Research* 53: 1-42.

[1](#) I use this abbreviation for 'if and only if'.

[2](#) Bonjour, pp.9-10.

[3](#) Kripke, p.48. We may set aside worlds where Bush does not exist – the name 'GWB' will have a null extension in such worlds, but that doesn't affect the present arguments.

[4](#) Kripke, pp.55-56.

[5](#) Kripke, p.77.

[6](#) Hence for any possible world  $w$ , ‘actually P’ is true in  $w$  iff P is true in @. (Davies and Humberstone.)

[7](#) Kripke, pp.100-105. But cf. Soames.

[8](#) Kripke, p.122. I also borrow Kripke’s “cat” example.

[9](#) Putnam, p.232.

[10](#) And not just due to rational failures on our part, but because the statement is not a priori knowable at all. For ease of exposition, I won’t explicitly repeat this in what follows, but my claims should be understood as applying to ideally rational agents who know all *a priori* truths.

[11](#) Cf. Jackson, p. 77.

[12](#) Cf. Sidelle, chp.2.

[13](#) Jackson, pp.77-78.

[14](#) Of course, anyone may grant that what language we’re speaking may depend on what world we’re in. I’m talking about a stronger phenomenon, viz. the opposite of “semantic neutrality” as defined in the next section.

[15](#) See, e.g., Chalmers (forthcoming).

[16](#) Cf. Chalmers (ms.) 4.6.

[17](#) *Ibid.* Technically, primary intensions are functions from *centered* worlds to extensions, but I will set aside this complication for now.

[18](#) Though it’s worth noting that rigid designators *can* be semantically neutral, so long as they are also epistemically rigid, i.e. picking out the same entity in all worlds considered as actual. Terms for abstract objects, like ‘seven’, might work this way, for example. Chalmers (ms.), 5.7.

[19](#) Again, the modal rationalist should strictly speak of ‘centered possible worlds’, but I set aside this complication for the purposes of our discussion.

[20](#) Chalmers (1999), section 3.

[21](#) Unless otherwise stated, I will always treat scenarios as being ‘considered as actual’. So here “falsify...” means “verify (not necessarily satisfy) the negation of...”

[22](#) Chalmers (2002), section 12.i.

[23](#) Roughly, we can associate scenarios with possible worlds considered as actual. There are various complications – e.g. ‘centering’ for indexicals – discussed in Chalmers (ms.) chp.5, which we needn’t get into here.

[24](#) Chalmers (1999).

[25](#) Proper names usually aren’t semantically neutral – so would be disqualified from featuring in conceivability-possibility inferences as explained in Chp.1 – but ‘God’ as used here is meant to be super-rigid, referring to the same entity in all worlds *and scenarios* in which it exists. (Cf. ‘Tully’, which refers to someone other than Cicero in some scenarios.) In any case, a similar argument could be run using the more obviously semantically neutral claim, ‘a necessary being exists’; but I use ‘God’ for ease of exposition.

[26](#) Note that the following is a reconstruction of Yablo’s argument.

[27](#) This way of formalizing the premise serves to simplify the logic of the argument. Purists might prefer to begin with ( $\blacklozenge G \ \& \ \blacklozenge \sim G$ ), but since God is taken to be a necessary being,  $G$  is understood to be equivalent to  $\Box G$ , and  $\sim G$  likewise equivalent to both  $\sim \Box G$  and  $\Box \sim G$ , as should be intuitively clear from the accompanying discussion. (Note that the equivalence of  $\sim \Box G$  and  $\Box \sim G$  follows from the impossibility of there existing a necessary being that does not exist in all worlds. Plainly such a being would then be contingent, not necessary, hence contradicting the initial specification.)

[28](#) Contrast the negation of  $S5$ , which is equivalent to  $\Diamond \Box P \ \& \ \sim \Box P$ . That is, “there is a possible world  $w$  according to which  $P$  is true at all possible worlds; but it is not the case that  $P$  is true at all worlds.” The possible modal space posited by  $w$  would thus differ from actual modal space.

[29](#) Howell; see also Chalmers (1999) section 3.3.

[30](#) Chalmers (1999) section 3.3.

[31](#) Chalmers (2002).

[32](#) Subject to the 2-D semantical complications discussed in Chapter One.

[33](#) Granted, it is independent of any actual minds. But insofar as a class of facts is defined in terms of (even merely possible) idealized minds, it remains less than *fully* mind-independent. Cf. Sturgeon.

[34](#) Others might split the philosophical duties more evenly, perhaps granting that entailment and semantic values go with conceptual space, whilst the more “metaphysical” issues like supervenience, properties, etc., go with metaphysical possibility. Cf. Edgington. But if ours was the *only* metaphysically possible world, in the ‘realist’ sense, that shouldn’t lead us to think that all actually coincident properties are identical, or that every fact trivially supervenes on every other one. The “subjunctive possibility” yielded by the 2-D framework applied to epistemic space, would seem much better at filling the theoretical role of so-called “metaphysical possibility”. See also Chalmers (1999) 3.5: “A narrower class of worlds is no help in making sense of these notions; breaking the tie between conceivability and possibility breaks the tie between rationality and modality.”

[35](#) Cf. Van Inwagen, p.71: “It hardly follows that, because a certain thing cannot be proved to be impossible by a certain method, it is therefore possible in any sense of ‘possible’ whatever.” What I here call “content-based” kinds of possibility are, for Van Inwagen, mere pseudo-possibilities.

[36](#) Other times it may be understood as ‘compossibility’ with the laws of nature, in which case something that is metaphysically impossible would be considered “nomologically impossible” even if the laws of nature alone provide no grounds for ruling it out.

[37](#) Chalmers (1999), section 3.5.

[38](#) Similarly, there might be once-open futures that are no longer accessible to us given our present circumstances.

[39](#) Edgington, p.6, calls the timeless version “absolute metaphysical possibility”; her time-relative notion of “metaphysical possibility” appears to be what I here call “dynamic possibility”.

[40](#) Cf. Fine, section 3, on “the peculiarly modal force of truths that are naturally necessary yet metaphysically contingent.”

[41](#) Chalmers has suggested to me that the converse may hold, i.e. that the *laws* are defined in terms of what is necessary in this primitive sense. But this risks undermining what scientists normally take to be natural laws. We may prefer to hold that the deep explanatory regularities that scientists discover are “natural laws” – even if there was a chance, at the beginning of time, that others might have taken their place. But this may be a merely terminology dispute in any case. The crucial point is that the modal properties of the regularities we typically assume to be laws – e.g. concerning gravitational attraction, etc. – are not directly fixed by the primitive modal concept I’m pointing to.

[42](#) Cf. Stalnaker.

[43](#) That is, a fact about *what there is* rather than *what could be*. We may think that there are necessarily uninstantiated properties, e.g. ‘being a round square’. Perhaps there are some maximal world-properties that are like this.

[44](#) It’s worth noting that my above attempts to motivate these principles in no way depend on which world is actual, suggesting that they will be a priori if they are justified at all – just as modal rationalism itself demands.