

ANTINOMIES OF PURE CONCEPTION: AN ESSAY ON
CONCEIVABILITY ARGUMENTS

by
Stephen D. Harrop

A thesis submitted to Johns Hopkins University in
conformity with the requirements for the degree of
Bachelor of Arts in Philosophy with Honors

Baltimore, Maryland
March, 2016

© 2016 Stephen Harrop
All Rights Reserved

Abstract

Historically, it has been thought that there is some connection, whether structural or evidential, between conceivability (in a suitable sense) and possibility. Recent years have seen attempts to resurrect this connection, both by means of arguing that conceivability is a *guide* to possibility and by arguing that conceivability *entails* possibility (I call this the “royal road”). This paper concerns the latter connection. I argue that such principles ought, if they are to be thought fruitful at all, to be formulated in terms of an idealized conceiver: something is conceivable *iff* an ideal conceiver could conceive of it. I continue by arguing that, construed this way, the royal road’s converse–inconceivability entails impossibility–overgenerates impossibility claims. That is, it returns as inconceivable things which we have good reason to believe are possible (in some cases, actual). I conclude that this gives us reason to reject the dual, which in turn gives us reason to doubt the “royal road.”

Contents

Abstract	1
Acknowledgments	5
Chapter 0. Preliminaries	7
Chapter 1. The Royal Road	13
0. How might conceivability entail possibility?	15
1. Agent-relative vs. Agent-invariant	18
2. Empiricist vs. Rationalist	30
2.1. The empiricist account	31
2.2. The rationalist account	37
3. Recap	43
Chapter 2. Antomies of Pure Conception	47
0. Modal monism	48
0.1. Expansions of (1)	48
0.2. Expansions of (2)	50
1. Against modal monism	52
1.1. A posteriori necessities	53
2. <i>Inconceivable!</i>	58
2.1. Does $\text{Inconceivability}_P$ Entail Impossibility?	58
2.2. Does $\text{Inconceivability}_O$ Entail Impossibility?	62
2.3. Does Ideal Negative Inconceivability entail Impossibility?	68
3. The upshot	85
Closing Remarks	95
Bibliography	99

Acknowledgments

Big things have small beginnings; small things, smaller still. Whatever (small) novel ideas find their expression in this thesis grew in the telling. Its germ arose in an exchange (between myself and graduate student David Williams, if I recall correctly) during a 2014 seminar on the history of philosophical notions of possibility. Its first leaves sprouted as a term paper written for the same class. And it began to truly bear fruit in the reflection following the paper. Whatever finished product follows is the result of both reflection and engagement with my teachers and peers.

Not a few of these deserve my thanks. Foremost among these is Yitzhak Melamed, who endured my questions and challenges—and, if I am perfectly honest, frequent insouciance—with aplomb, understanding, and a readiness to answer just about any question, no matter how ineptly expressed or incendiary. He has been an exemplary advisor and intellectual mentor. Less directly academic but still substantive thanks are due to Nandi Theunissen, for her aid in my being able to submit this thesis in the first hand, as well as her readiness to help an undergraduate with pretensions to philosophizing to begin tentative steps towards making those pretensions realities.

Also deserving thanks are some of my fellow undergraduates and graduate students: David Williams, for prompting the initial thoughts which are here expressed in (I hope) a slightly more polished fashion; Virgil DeMario, whose ready ear to my frequently-incoherent ramblings forced me to make them as precise as possible; and Jason Paone (formerly of UT Austin, now located at Duke Divinity), for comments on the original paper and for general (and much-needed) notes on style and academic convention.

Of course, personal acknowledgments are due to: my mother, Karen Harrop, for her devotion to ensuring that I have the skills and opportunity to obtain the best education possible, and for her tireless commitment to encouraging her children to enjoy learning both in class and out; and to my (eventual) bride-to-be Susanna Kellogg, for her tolerance of my brown studies during the writing of this thesis, and for her unflagging support of my academic pretensions and interests. I owe them both irrefragable debts.

Finally, *ITG*.

CHAPTER 0

Preliminaries

Throughout the history of philosophy, conceivability has been thought to stand in a very close relation to possibility. Thomas Reid reports that “[t]his opinion has been held by philosophers for more than a hundred years without contradiction or dissent, as far as I know.”¹ He gives a list of celebrated philosophers to whom such a belief can be ascribed; amongst the names are Samuel Clarke, Christian Wolff, and David Hume. Similarly, Kant writes in “The Only Possible Argument”:

[E]verything possible is something which can be thought, and the logical relation pertains to it in accordance with the principle of contradiction.²

Today, well over two centuries after these works were written, there has been considerable attention paid to the connection between epistemic modality–possibility in the sense of “true for all I know,” necessity in terms of the a priori–and metaphysical modality–which is somewhat more difficult to characterize in terms of a definition. The usage in modern analytic philosophy seems to originate primarily in Kripke:

We ask whether something might have been true, or might have been false. Well, if something is false,

¹Reid (1941), p. 256.

²Kant (2003), p. 123.

it's obviously not necessarily true. If it is true, might it have been otherwise? Is it possible that, in this respect, the world should have been different from the way it is? If the answer is 'no', then this fact about the world is a necessary one. If the answer is 'yes', then this fact about the world is a contingent one. *This in and of itself has nothing to do with anyone's knowledge of anything.*³

Kripke draws a sharp distinction between the a priori and the necessary, and the impulse of many philosophers has been to follow him in this. Not everyone has been inclined to agree with this assessment, and so the discussion continues today. One of the more interesting features of the debate has been the examination of the connection between a thing's being conceivable—for a suitable notion of "conceivable", of which more later—and a thing's being possible. Those who think there's such a connection take—roughly—one of two views. On one side, there are those who think that conceivability is a guide (though by no means a *certain* guide) to possibility. On the other, there are those who think that conceivability in some sense *entails* possibility. The former idea, while interesting in its own right, is not the topic of this essay. Rather, we will be taking up the latter notion—the question of whether conceivability entails possibility. This I will call the "royal road" (henceforth RR). The converse of RR is the thesis that inconceivability entails impossibility. The combination of RR and its converse yields what Vaidya (2008) calls "modal monism":

³Kripke (1980), p. 36. Emphasis mine.

[Modal monism] is a metaphysical thesis about the nature of metaphysical modality and logical modality. It maintains that although metaphysical modality is notionally distinct from logical modality, the two modalities are nevertheless extensionally equivalent.⁴⁵

Before we go on, some clarification is in order. Some philosophers maintain that we achieve modal knowledge via counterfactual reasoning. For instance, Timothy Williamson holds that knowledge of metaphysical possibility is a special case of counterfactual knowledge.⁶ This is not the view I will be examining in the following pages. Rather, as stated above, I will be interested in accounts of modal knowledge which turn on a thing's being *conceivable*. Consequently, when I speak of "modal monism," I have in mind a version of the latter whose account of epistemic possibility relies on conceivability as the (or a major) source of modal knowledge. For the purpose of this thesis, then, we will take both "modal monism" and "epistemic possibility/modality" to carry their conceivability-related meanings.

Returning to Reid for a moment, it is interesting to note that he did not think conceivability was a guide to possibility:

There remains another mistake concerning conception which deserves to be noticed. It is – That our conception of things is a test of their possibility, so

⁴Vaidya (2008), pp. 192–193.

⁵This is sometimes associated with certain interpretations of two-dimensional semantics. For a fuller exposition of that kind of modal monism, see Chalmers (2004), especially §3.9–3.11.

⁶See, e.g., Williamson (2007). for details, particularly §3.

that, what we can distinctly conceive, we may conclude to be possible; and of what is impossible, we can have no conception.⁷

Continuing:

[One of my reasons for believing this an error is]: Whatever is said to be possible or impossible is expressed by a proposition. Now, what is it to conceive a proposition? I think it is no more than to understand its meaning...I am persuaded that I understand as distinctly the meaning of this proposition, *Any two sides of a triangle are together equal to the third*, as of this – *Any two sides of a triangle are together greater than the third*; yet the first of these is impossible.⁸

I share much the same concern.

Broadly following Reid, then, the general thrust of this thesis goes as follows. First, under a suitable definition of inconceivable, there are inconceivable worlds. (This sounds more dramatic than it really is, as will become plain) Moreover, there are inconceivable possible worlds. Still more outrageously, the actual world is an inconceivable world. Consequently, modal monism is false. Further—or so the argument goes—this gives us good reason to suppose that it is unlikely that conceivability entails possibility.

The bulk of the first chapter is devoted to constructing adequate and reasonably general notions of conceivability. This, in turn, helps us to adequately formulate RR properly and specifically. I consider two broad categories of conceivability: agent-relative vs. agent-invariant and rationalist vs. empiricist. The agent-invariant

⁷Reid (1941), p. 256

⁸ibid., p. 259.

is cashed out in terms of what an “ideal conceiver” might be able to find conceivable. The rationalist version determines whether or not a thing is conceivable based primarily on contradiction-detection. The empiricist version cashes out conceivability in terms of what an ideal conceiver might be able to mentally, and sensorily or quasi-sensorily, represent.

The second chapter is devoted to examining RR’s dual. The general question of the chapter is: does inconceivability imply impossibility? I argue that it doesn’t, based primarily on the reasoning that the relevant notions of conceivability entail that certain things which we have good reason to believe are possible—indeed, actual in some cases—are in fact impossible. This, I claim, gives us good reason to reject the idea that inconceivability entails possibility. Furthermore, though with somewhat lesser strength, I argue that this in turn gives us good reason to reject the idea that conceivability entails possibility.

Before going on, allow me a brief detour concerning singular pronouns. Throughout most of the history of (modern) English, as far as I can tell, the pronoun “he” was equally well usable as masculine nominative and as neuter nominative (ditto for its accusative, genitive, and dative cases). In current writing, this practice has fallen out of favor. Authors of both sexes have tended to opt either for “she” or “they” as the singular neuter (ditto, again, for the other cases). I find the former practice peculiar; and as for the latter, I cannot bring myself to believe that it truly is good orthographical practice to transform a plural neuter into a singular one.

Consequently, I shall default to the following rule of usage: the author shall use, as the singular neuter, that pronouns whose gender corresponds to his own. In this paper, I will use “he” (and the corresponding variations for the appropriate cases) in the indicated way.

With that out of the way, we turn to the following question: “When we say that conceivability entails possibility, what do we mean by ‘conceivability’?”

CHAPTER 1

The Royal Road

“The thought contains the possibility of the state of affairs which it thinks. What is thinkable is also possible.”

– Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, §3.02

Suppose Richard and I are having a conversation. During the course of the dialog, I say “suppose that extraterrestrials had visited Earth. Then—” “Hold on,” Richard responds, “how have you got extraterrestrials in your thought experiment?” Upon reflection, I say to him, “well, I seem to be able to conceive of a situation in which extraterrestrials visit Earth; consequently, it’s possible that they have or will visit Earth.” We then move on and continue in our discussion.

But is my reasoning good? That is, does the fact that I can conceive of P—where P is an entity, state of affairs, or what-have-you—give me warrant to conclude that P is in fact possible? It is this question that I will be examining in what follows.

There are at least two ways in which I might be construing the line of reasoning presented in the hypothetical dialog. One is to say that that conceivability is a reliable guide to possibility, but does not actually entail it. One could think this for systematic reasons—because our abilities to conceive of P are grounded in our concepts and our concepts themselves are grounded in what is actual

(and therefore possible)—or one could give a heuristic or pragmatic justification-conceivability produces more “yes” answers than “no” answers when used to answer the question “is P possible?” This is roughly the route taken by Stephen Yablo.

Alternately, one might say that conceivability entails possibility, as opposed to merely providing a usually-reliable guide to modal truths. On this view conceivability is thought to stand in some sort of structural or logical relation to possibility, such as logical entailment. Stronger versions might hold that (some suitable version of) conceivability is, in fact, equivalent possibility. The former view has been defended by David Chalmers, the latter by both Frank Jackson and, in some places, Chalmers.

A related question is how we ought to construe conceivability. One way, which appears as early as the British empiricists (in particular Reid), is to understand conceivability as a sort of quasi-perceptual experience, a sort of mental image.¹ To conceive of P, on this model, is to have a sort of mental representation of P. Another way is to take conceivability as less of a sensory-based mental representation and more a matter of examination of concepts, weighed against each other to determine compatibility or incompatibility, with no necessary quasi-sensory component. As such, these views are usually rationalist or a priori in nature. Something like the former view is defended variously by Yablo and by Peter Menzies; Chalmers has defended something like the latter.

¹Though it should be noted that this was not necessarily the common use throughout the history of philosophy.

This section is concerned primarily with examining various entailment relations which might obtain conceivability and possibility, with the evidentiary views taking second fiddle. We will consider, however, how we might go about converting a reliabilist or evidentiary thesis of the conceivability-possibility view to an entailment thesis. We shall also be interested in whether any relation of this kind must be a thinker-dependent account of modal terms—that is, whether we should not only treat conceivability as entailing possibility, but also identify it with possibility (and thus impossibility and necessity with their conceivabilistic counterparts).

0. How might conceivability entail possibility?

There are a variety of ways into which one might split regarding the role of conceivability in modal thought. For instance, Sara Worley claims that “whether or not something is conceivable for a thinker depends on what the thinker knows or believes, or what concepts or modes of presentation he has available or is using to think about the situation.”² Another version is presented by Heir Geirsson: “to be conceivable is to be true in a possible world, [. . .] where the possibility is determined by conceptual coherence or incoherence, and thought in terms of an ideal conceiver.”³ Similarly, Menzies writes, “[I]t is possible that p if and only if an ideal conceiver could conceive that p.”⁴

²Worley (2003), p.17.

³Geirsson (2005), p. 290.

⁴Menzies (1998), p. 269.

Roca-Royes (2011) calls these epistemic and non-epistemic senses, respectively.⁵ For our purposes, we'll refer to these senses of conceivability as agent-relative and agent-invariant, for reasons that should become clear shortly.

Another way that we might divide up the concept of conceivability is between what Murphy (2005) calls the empiricist and rationalist accounts. The empiricist account “[M]akes imagination central to conceivability and inconceivability[...]. Thought of this way, conceivability is a two-step cognitive process: a cognizer imagines a situation (or situations) that she takes to verify p; and this causes her to believe that it is possible that p.”⁶ On the rationalist account, “[C]onceivability and inconceivability do not involve any images or sensory-like states. Instead, they involve contradiction detection[...]. [F]inding p inconceivable involves detecting a contradiction in p and subsequently believing that p is impossible.”⁷ Murphy’s terminology is both useful and illustrative, as it accurately captures the salient way in which the two accounts differ: one view takes conceivability to be an irreducibly sensory or quasi-sensory action, while the other is concerned mostly with relations between concepts, without reference to a particular sort of mental representation. For these reasons, I will follow Murphy in this usage.

⁵See Roca-Royes (2011a) for further details.

⁶Murphy (2006), p. 196.

⁷Murphy (2006), p. 197.

There are still further ways in which one might break down the concept of conceivability. One might be something like the following: it is conceivable for X that P if X holds no belief inconsistent, upon reflection, with believing that P. This is related to the rationalist version in a deep way, but is not identical to it. One need not detect that there is no contradiction in believing that P in order to find P conceivable in the sense just given. This is the rawest sort of epistemic possibility—something is conceivable for me just in case it might be true “for all I know.”

On the face of it, it’s not obvious which of these versions of conceivability we should use to formulate RR. Depending on which version of conceivability we end up using, we might in fact have a different sort of conceivability-to-possibility principle. Since we are interested in *structural* or *logical* relations between conceivability and possibility, it might be useful to put the putative relation in a formal manner. On first blush, we might write it as something like the following:

$$(1) \quad C\phi \rightarrow \diamond\phi,$$

where $C\phi$ denotes “it is conceivable that ϕ .”

But ought we to take C as primitive? Or ought we break it down further, to get a better idea of the lifting it’s supposed to be doing? If we think of “conceiving” as a primitive activity, it’s hard to know how we might actually provide an analytical account of *how* this relationship might work. Consequently, our operative assumption will be that it can in fact be analyzed further. We might start with

the following:

$$(2) \quad Ca\phi =_{df} Ja\Diamond\phi$$

where Ja is taken to mean “ a judges that”. A great deal hinges around the different definitions we might give to the phrase “ a judges that”. In the spirit of greater ecumenism, we’ll take “ a judges that” to mean “ a carries out whatever process or activity that we take to be constitutive of conceiving.” Hardly a concise interpretation, but that comes at the cost of generality. We can therefore consider different versions of our entailment relation in parallel with the various sorts of conceivability we wish to discuss.

So, with the symbol defined thusly, we can write:

$$(3) \quad Ja\Diamond\phi \rightarrow \Diamond\phi$$

A useful—though possibly beguiling if taken too literally—way to think of this is as a sort of elimination rule. If we’re able to reach a point in our reasoning where we’ve got a statement of the form “agent a judges that ϕ is possible,” where ϕ is some appropriate proposition, entity, or state of affairs, then we’re entitled to eliminate the “agent a judges that” bit and simply conclude “ ϕ is possible.”

We can now consider how precisely this inference might interact with the different sorts of conceivability given above.

1. Agent-relative vs. Agent-invariant

First, let’s examine the agent-relative sense of conceivability. In this case, our entailment would be “if it is conceivable to a that ϕ ,

then it is possible that ϕ ." Here, we treat a as a placeholder for whichever agent we want to consider. Now, as it stands, it turns out that this is not truth preserving across different agents and different ϕ . For instance, suppose that our agent is a standard-issue 6th grader, and suppose that ϕ is the proposition "all infinite sets are of the same size."⁸ Now suppose that the 6th-grader indeed finds it conceivable—in the sense of making proper and non-contradictory reference to the agent's modes of representation, beliefs, and so forth—that all infinite sets are of equal cardinality.⁹ But, of course, that is in fact not the case—for instance, the cardinality of the continuum is strictly greater than the cardinality of the natural numbers. So we have a true antecedent and a false consequent; thus, the inference is false.

This failure might predispose the reader to despair of the usefulness of this notion.¹⁰ Such despair would be premature. For, remember, proponents of agent-relative conceivability usually relativize the inference to: "it is conceivable for a that ϕ " implies "it is possible for ϕ that x ." That is, the antecedent is itself modified to take on the meaning of, roughly, "it appears to a that possibly x ."

⁸This assumes the 6th grader has had some set theory; I myself did at about that time, but I do not know whether or not that forms part of the standard curriculum.

⁹Of course, the 6th grader will likely just take this to mean "all infinite sets are the same size." Since I have some reservations about the use of "size" in this context, I will use the more technical and better-defined (in the infinite case) term "cardinality."

¹⁰George Bealer, for instance, writes: "The mere fact that I tried, but happened to fail, to conceive that p is not a good guide to what is in principle possible in this regard for any being whatsoever; maybe I am just not sharp enough...The moral is simple: in the matter of evidence for possibility and impossibility, talk of conceivability and inconceivability is an idle complication that only breeds confusion." See [Bealer \(2002\)](#), p. 76.

So returning to our case of the 6th grader, since he does in fact find it conceivable that all infinite sets are of the same size, I see no great objection to concluding that “it appears to our 6th grader that it is possible that all infinite sets are of the same size.” As a consequence of this, we might think to modify our entailment to the following:

$$(4) \quad Ca\phi \rightarrow \diamond_a\phi$$

where $\diamond_a\phi$ is to be understood as standing for “it appears to a that possibly ϕ .”

So far, so good. How do we handle the agent-invariant case? At the outset we should note a crucial difference between these cases which the terminology being used might have obscured. The agent-relative rendering is primarily concerned with what a particular agent may reasonably conclude is possible, based upon what he determines is conceivable—thus, any possibility it might obtain is, most likely, epistemic. The agent-invariant version, on the other hand, could in principle be free from the sorts of pitfalls which led us astray in the sense of the 6th grader. This version, whatever its demerits, is precisely *not* intimately wedded to talking about what *this* particular agent might find conceivable. Instead, if it is to determine that something is conceivable, this will have to hold true independently of the foibles of any particular agent or conceiver. This gives it—at least, on the face of it—a much better shot at getting to a thicker, more metaphysical notion of possibility from some conceivability statement.

One way we might achieve this sort of neutrality is to use Menzies' conceit of the ideal conceiver. This pretty naturally requires us to modify the previous entailment relation, which, recall, was

$$(5) \quad J a \diamond \phi \rightarrow \diamond_a \phi$$

The problem with this version—or rather what forced us to modify the original—was the fact that we treated “*a*” as free, or at least interchangeable. This in turn kept the entailment from being truth-preserving, since what counts as conceivable will vary greatly from person to person. Ideally, this version won't suffer that problem. The entire point of speaking of an ideal conceiver is precisely that it should be free of these pitfalls. At a first approximation, we might modify the above entailment to:

$$(6) \quad J_I \diamond \phi \rightarrow \diamond \phi$$

where *I* is an ideal conceiver. The success of this scheme pretty evidently depends on how we're cashing out the notion of an ideal conceiver. We might describe *I* as follows:

DEFINITION 1. (*Ideal Conceiver*) A conceiver *I* is called ideal if:

- (1) *He possesses ideally excellent mental and cognitive capacities.*
(*Ideality*)
- (2) *He possesses only true beliefs.* (*Doxastic soundness*)
- (3) *He is capable of completing any rational or extrapolative process from his basic belief set in a finite number of steps.* (*Finiteness*)

The need for (1) should be pretty clear—if we want an ideal conceiver, certainly his mental and cognitive capacities should be as finely tuned as one might desire. (2) is pretty obviously necessary as well; as we saw in the case of the 6th grader, possession of false beliefs can result in the entailment not holding. Though the entailment would trivially hold in the case of the antecedent and the consequent both being false, or vacuously hold in the case of the consequent being true and the antecedent false, neither of these are of any good epistemic or metaphysical use. For our purposes, these cases are boring. The non-boring case, then, is the one in which both antecedent and consequent are true. Additionally, since \mathcal{I} 's an ideal conceiver, he should ideally (pun intended) have access only to good information. For our purposes, “good” should be taken to mean “maximally accurate.”

It's possible that (3) is not strictly necessary, but it seems like a good stipulation nonetheless. In some situations, determining whether or not a given proposition, state of affairs, or entity is in fact conceivable (specifically, in the case where the conceivable is identified as whatever is determinable a priori) would require \mathcal{I} to compare various elements of his basic epistemic set with the concept in question to see whether any inconsistency arises. If such a process never terminates, we reach an obvious difficulty: \mathcal{I} will never determine whether a given thing is conceivable, even though it very well may be possible. Such a conceiver could hardly be called “ideal”!

To give an example (which will be somewhat misleading if taken too literally): suppose that our ideal conceiver wants to determine if there is a highest prime number. And suppose he elects to do so simply by computing the prime factorization of every natural number. This process will never terminate, since to determine the conceivability of this proposition in this way would require a countably infinite (at least) number of steps. In this case there is, of course, a very simple way to arrive at the truth of whether the given proposition is conceivable. And that would be to produce a proof of Euclid's theorem that is itself checkable in a finite number of steps. The point which we are attempting to illustrate is that without some restrictions, some processes of conceiving will never terminate, even though the thing conceived of in fact has a definite modal status. This motivates (3).

This will also place restraints on the person who holds that to be conceivable is to be determinable a priori, in some sense—that is, in order to determine whether or not something is conceivable, \mathcal{I} has to check its coherence with his basis belief set. Call the ideal conceiver's basis belief set B , and the set of all propositions compatible with the propositions expressed by the conceiver's doxastic attitudes CB . Then, if (3) is to be satisfied, CB must (accepting the Church-Turing thesis) be at least recursively enumerable, since otherwise there is no way to determine whether or not a given proposition is a member of CB (which is another way of saying "is compatible with every element of B "). As far as I can tell, this holds

only for the versions of conceivability which require a priori coherence (of which, more later). So it looks like we'll have to Chisolm away at (2) a bit, to something like the following:

DEFINITION 2. (Doxastic soundness and semi-decidability) *I possesses only true beliefs, and the set of all propositions consistent with those beliefs is at least semi-decidable.*

A potential objection to this definition suggests itself here. Given some set of true beliefs B_i , there must be a corresponding set of propositions consistent with it, CB_i . If we want to be able to determine what's in CB_i , however, it seems like CB_i needs to be at least semi-decidable. This implies that, if we were to speak of a *maximally* complete set of true beliefs—that is, one which just contains any and all true beliefs not involving personalized true beliefs (e.g., indexical ones: “That baby is mine”)—then the corresponding set CB will have to be decidable. This is an odd consequence, since it seems to give an upper limit—or at least a limiting principle—to how many true beliefs there might be. Consequently, one might say, how many things there are in heaven and earth would be *determined* by what is dreamt of in our philosophies. Or, more minimally, if one has a set of all true beliefs but not necessarily a maximally complete one, there still couldn't be any true propositions which don't cohere with them (that is, which contradict or are inconsistent with them). The idea, then, is that this is a monumentally ambitious thesis, which gives us reason to be hesitant in its acceptance. This is an interesting objection, but not, I think, a very good one, because it amounts only

to saying “you’re proposing rationalism!” That is as it may be. But the entire notion of the identification of the conceivable with the a priori was precisely a proposing of a modal rationalism, and one can’t very well object to a philosophical thesis on the grounds that, if it were true, it would be true.

Another way we might solve this problem is by stipulating that, in addition to (1), our ideal conceiver has amongst his cognitive apparatus something like a “black box.” That is, for a given computation that cannot be carried out in a finite number of steps, the ideal conceiver is able to determine, in a single or very small number of finite steps, its modal status. Methodologically, however, this is a poor move, in part because it amounts to a fudge, but also because part of this whole exercise is to determine whether or not modal truths can be arrived at by human reasoners, even if they be humans of exalted capability. Certainly, humans can intuit the answers to some questions before they can arrive at a strict demonstration of it.¹¹ But this seems unlikely to be the sort of thing that could be thought to be in common to humans or reasoners as such, and it certainly does not seem like the kind of thing which might be thought of as a process of *computation*. So: no oracles.

This leads up naturally to our next point: in principle the ideal conceiver should possess only maximally fine-tuned human capabilities. If we’re interested in knowing whether our intuitions about

¹¹A canonical example is the Riemann hypothesis. It is my sense that the majority of complex analysts tend to think it’s true, and to date all of the poles of $\zeta(s)$ which would qualify do in fact line up on $\text{Re}\{z\} = \frac{1}{2}$, but so far no proof has been forthcoming.

modality or our conceiving abilities entail modal truths—or even the weaker proposition that they are a usually reliable guide to modal truths—then we are not interested in what some non-human conceiver—that is, a conceiver whose has a completely different set of mental faculties than we do—might be able to tell us conceivability, unless this ideal conceiver was able to arrive at the modal truths he imparts to us in a manner which only made use of distinctively human faculties and beliefs which humans could, under ideal circumstances, obtain. And at that point, there’s no need to posit a non-human conceiver when simply an ideal human one will do.

So we arrive at the following amended definition:

DEFINITION 3. (Ideal Conceiver, revised) *A conceiver \mathcal{I} is called ideal iff*

- (1) *He possesses ideally excellent mental and cognitive capacities. (Ideality)*
- (2) *He possesses only true beliefs. (Epistemic Soundness)*
- (3) *He is capable of completing any rational or extrapolative process from his basic belief set in a finite number of steps. (Finiteness)*
- (4) *The set of all propositions consistent with his basic belief set is semi-decidable. (Checkability)*

So, with this in place, we can reduce our entailment relation to:

$$(7) \quad J_{\mathcal{I}}\diamond\phi \rightarrow \diamond\phi$$

where we’ve replaced the $J_a\phi$ with $J_{\mathcal{I}}$, to indicate “ \mathcal{I} judges that...” etc. It might seem that this formulation of the non-epistemic

version in fact makes it into an epistemic version yet again. And this is not entirely inaccurate. But the chief drawback with the original epistemic version was that it offered a good way to determine epistemic, but not *metaphysical* modality. And, while the agent-invariant notion is dependent in some sense on the basic belief set of \mathcal{I} —and is consequently and in this case an epistemic formulation—the determinations of \mathcal{I} , as formulated, will, as intended, be truth-tracking in the proper way.

At this point, having spelled out what our ideal conceiver would look like, we could cash out the significance of this conceit in one of three ways. First, it could be cashed out as what any agent would be able to conceive of if he were “uplifted” and acquired all the relevant characteristics of the ideal conceiver which he did not already possess. Second, it could be cashed out in a more a priori sense, with the ideal conceiver being a useful heuristic that simply refers to what is computable or deducible under ideal circumstances, without any particular commitment to whether or not the ideal conceiver actually exists. Third, one could avoid the “uplift” method, but still think that the talk of “ideal conceiving” is not a mere useful fiction.

The first method is for all intents and purposes a way to move from the agent-relative case to the agent-invariant case. Consider again our intrepid 6th grader. He was not able to conceive of there being different sizes of infinity, and consequently our conditional $\diamond_a\phi \rightarrow \diamond\phi$ failed, since it had a true antecedent and a false consequent. But most 6th graders grow up. At one time I had no notion

of different sizes of infinity, and now I do. I can think back, and say something like the following: had I known then what I knew know, I would have been able to conceive of different large cardinals in the appropriate way. We engage—or at any rate I engage—in this pattern of reasoning all the time, whether with regard to important matters (“had he known of the 2008 housing bubble, he would not have invested in real estate”) or with regard to trivial ones (“had I known that she intended to use the Shepherd’s Check, I would have moved my pawn differently”), and we take ourselves not to be talking nonsense in these cases.

Motivated by this, we can try and cast the inference relation informally as the following:

DEFINITION 4. (Ideal Conceiver Entailment) *For an agent a and an entity ϕ , if a were to possess 1-4 and could conceive of ϕ , then possibly ϕ .*¹²

I refrain from setting this out formally because I think the structure of general RR would remain more or less unchanged from the general agent-invariant case. The only difference in this case is that we are seeking out what particular interpretation we want to give to the content of the inference.

The reader might notice that, though the agent in question is being “uplifted,” his base belief set remains unchanged. More to the point, while the agent’s base belief set might conceivably (no pun

¹²Sidenote: one alternate, formalized version of this sort of implication which I think worth exploring is $C_I\phi \diamond\rightarrow \phi$. Since this would take us into accounts of modal knowledge grounded in counterfactuals—and we’ve bracketed those out for the sake of discussion—we will stick to the simple material conditional.

intended) be pared down a bit—since most agents we would care to work with will have some false beliefs—it will not be expanded. This should be unsurprising, since the process is, so to speak, one of refinement rather than total expansion. This has the potential to render the account somewhat agent-relative, absent some other constraint placed on the uplift process.

Were we to endorse the second method, we would be committed to removing the conceiver from the picture entirely. Or, at the very least, we would be adopting the view that such an agent is not an indispensable (or perhaps even unnecessary) part of our system. Since some of the properties of our ideal conceiver do in fact make reference to agent-specific properties (“excellent cognitive and mental faculties”), we would have to adopt some sort of revisionist attitude towards it. Let’s say, for instance, that we think that 1)-4) represent (implicitly, admittedly) a sort of multiply realizable process rather than the entity in which the process is realized. We might then want to rephrase 1) to something like the following:

DEFINITION 5. (Ideality, revised) *The \mathcal{I} -process contains no faulty or non-truth preserving computational or comparative mechanisms or algorithms.*

(2)-(4) work pretty much equally well in either case, since up to relabeling they don’t require any consequential changes.

This process seems to make an interesting concession. Recall that we are, at bottom, interested in precisely how our armchair

speculations about modality might be a good guide to true metaphysical modality. As a result, if we're to spell out this "ideal conceiver" in terms of an abstract, multiply-realizable process or entity, then it will have to be one which could in principle be carried out by a human agent. Otherwise, this would be an interesting but academic exercise about what could, in principle, count as a good account of conceivability. And while such an exercise is in itself quite interesting—of philosophical value, even—that's not what we're presently aimed at. Consequently, the computational or comparative mechanisms or algorithms made use of in this process have to be ones which a given agent presently possesses and/or employs, however imperfectly.¹³

2. Empiricist vs. Rationalist

It should be noted that the ways in which we parsed out conceivability at first are not properly symmetric. The agent-invariant and agent-relative senses are in a sense a first-order thesis about conceivability—they involve, primarily, who does the conceiving. On the other hand, there are a variety of independent, second-order theses about what conceivability is. The ones we will examine in a moment are primarily concerned not with the agent doing the conceiving, but rather with what precisely the "*a* judges ϕ " clause in our formulation of RR involves. The divide between the empiricist and rationalist account, naturally, turns on what conceiving entails.

¹³There is another assumption at work here: conceivability, in the sense given above, is either computational or algorithmic in nature (or at least arbitrarily well-modeled in those ways). The implications of this view are potentially interesting, but I will not examine them here.

Roughly speaking, the empiricist account holds that conceivability is (at least primarily) a phenomenal act, involving sensory or quasi-sensory capacities and activities. The rationalist account, on the other hand, does not (necessarily) involve any sort of sensory element, but rather is concerned with determining coherence, or lack of contradiction.

2.1. The empiricist account. The seeds of the empiricist position go at the very least back to the Thomas Reid. He writes:

Conceiving, as well as projecting or resolving, are what the schoolmen called *immanent* acts of the mind, which produce nothing beyond themselves...That this is the common sense of men who are untutored by philosophy appears in their language. If one ignorant of the language should ask, What is meant by conceiving a thing? we should very naturally answer that it is having an image of it in the mind—and perhaps we could not explain the word better. This shows that conception, and the image of the thing in the mind, are synonymous expressions. The image in the mind, therefore, is not the object of conception, nor is it any effect produced by conception as a cause. It is conception itself. That very mode of thinking which we call conception is by another name called an image in the mind¹⁴

In some current debates, this account of conceivability has gained some currency. For instance, Yablo writes:

I find *p* conceivable if I can imagine, not a situation in which I truly believe that *p*, but one of which I truly believe that *p*[...]Imagining can either

¹⁴Reid (1941), Essay IV, Chapter 1, §3 (p. 233).

be propositional–imagining that there is a tiger behind the curtain–or objectual– imagining the tiger itself.¹⁵

This he calls “philosophical conceivability.” In a similar vein, Jenkins (2010) writes that

[I]n at least some cases, the senses may ground modal knowledge by providing what I call epistemic grounding for our concepts, which concepts (help to) determine what we can and cannot conceive of, which in turn guides our modal beliefs. And[...]the fact that our concepts are grounded in this way is what accounts for the fact that conceivability — which I take to be a matter of what our concepts will allow us to conceive of — can be a guide to independent modal truth.¹⁶

Jenkins takes it that we can conceive of something if our concepts allow us to coherently represent a given entity (or state of affairs, or...). There’s a sort of two-level requirement going on. On one level resides the act of representing a given x —that is, “picturing” the given x in Reid’s sense. On another, our concepts, determined by our senses, which ground our powers of representation.

Following Yablo, we can offer a rough first definition:

DEFINITION 6. (Conceivability_E) *X can conceive of P iff X can imagine a situation of which X believes truly that P.*

Yablo is entirely correct that there are multiple senses under which we might understand “X can imagine.” First, we can consider what he calls *propositional* imagining. A chief differentiating

¹⁵Yablo (1993), p. 26.

¹⁶Jenkins (2010), p. 255.

feature is that this sense of imagining is not so much a phenomenal representation as it is a certain kind of propositional attitude. For instance, say that I have a mental state whose content can be expressed by “there is a 100 pound trout on the floor of the Milton S Eisenhower Library.” Sticking to the “propositional” case, the mental state would be (something like) having a propositional attitude—a belief, say—the content of which is accurately described by the English sentence “there is a 100 lb. trout on the floor of the Milton S Eisenhower library.”

We can attempt to collect this into a more concise definition:

DEFINITION 7. (Propositional Imagining) *A imagines_P P iff A has a propositional attitude whose content yields an accurate description (in some language) of P.*

Next is the “objectual imagining.” In Yablo’s case, this is to imagine the tiger itself. In our case, it would be imagining the giant trout itself, flopping about in the floor of the library. This sort of imagining involves, at least in some ways, a sensory or quasi-sensory presentation of the entity or state of affairs in question. This is a bit closer to the sense of “imagine” we (or at any rate I) tend to use in everyday conversation. It is also, note, approximately the one Reid offered above. Taking our cue from him, we can attempt a first-approximation at a definition:

DEFINITION 8. (Objectual imagining) *A objectually imagines (or “imagines_O”) P iff A has some sensory or quasi-sensory mental representation (or suitable analog of a mental representation) of P.*

Yablo notes—rightly—that in some cases imagining_O and imagining_P come hand in hand. There is also a question of the level of specificity which accompanies imagining_O:

To imagine an object as determinate is to imagine it as possessing the higher-order property stated, that of possessing a determinate property for each of its determinables. There is a world of difference, then, between imagining an object as determinate-as possessing determinates for each of its determinables—and determinately imagining it-specifying in each case what the underlying determinate is. What I have been urging is that objectual imagining is determinate in the first sense but not the second[...]What we are tempted to describe as imagining a more or less determinate situation, is better described as imagining a fully determinate situation whose determinate properties are left more or less unspecified.¹⁷

Given the degree of vagueness this seems to suggest, one wonders whether Yablo's account does the lifting he wants it to do. We can stipulate that these entities or states of affairs have some unspecified yet determinate properties or aspects, but this seems to give us only a partial guide to whether or not the complete thing in question is metaphysically possible.¹⁸ And this is a major *desideratum* of our

¹⁷Yablo (1993), p. 28.

¹⁸This essentially is the worry expressed in van Inwagen (1998): "Although it is in a sense trivial that to assert the possibility of P is to commit oneself to the possibility of a whole, coherent reality of which the truth of p is an integral part, examination of the attempts of philosophers to justify their modal convictions show that this triviality is rarely if ever an operative factor in these attempts. A philosopher will confidently say that a (naturally) purple cow is possible, but he or she will not in fact have devoted any thought to the question of whether there is a chemically possible purple pigment such that the coding for the structures that would be responsible for its production and its proper placement in a cow's coat could be coherently inserted into any DNA that was really cow DNA – or even "cow-like-thing-but-for-color" DNA." (p. 78).

current project. At this stage, it's hard to see how he might have a principled response to an objection such as the following: "these situations we're supposed to imagine may in fact have some deep-seated yet non-obvious flaw. Your account leaves certain amounts of vagueness in the description; but who knows, a priori, what filling in those dots would reveal? And if we *could* determine such details beforehand, why the vagueness in the first place?"¹⁹

One response to this sort of objection would be to say that the relevant respect of "imaginable" has been passed, and that that's all that matters. Take the example of the trout. I imagine_O the trout, flopping about on the floor of MSE. To be sure, I'm not having a mental representation of every aspect of the trout. I am not, for instance, imagining either a lake trout or a rainbow trout. Or perhaps I'm not really attentive to the differences between a 101 lb. trout and a 99 lb. trout. But certainly I'm imagining_O the trout as possessing *some* determinate properties—scales, for instance—and in that sense I can indeed conceive of the trout. If RR turns out to work, then one might simply start at the general conception of the trout. The first step would be to determine if any generic and specified properties of the trout are imaginable_O. Should they prove to be imaginable_O, then the imaginer can go further incrementally, testing each of the hypothetical, unspecified properties to see if they are imaginable_O.

¹⁹This is distinct from the problem of vague predicates. In typical cases of vagueness, the problem is that the predicates in question *do not admit* of specification. In our case, the problem is that the predicates *may* not have boundary cases, but are under-specified, leaving some degree of cognitive wiggle room.

In this respect, at least, imaginability_P enjoys a clear advantage over imaginability_O. While the details of a sensory or quasi-sensory mental attitude can be hard to determine or specify in all relevant respects, it is less hard to have a propositional attitude which is specified in as much detail as one might like. Consider again the case of the trout. Were we to use the imagine_P sense instead of the imagine_O sense, it would be fairly easy to simply include the other relevant properties, or at least a suitable and relatively determinate elliptical. The description could be something like “there is a 100 lb. trout, of ordinary features for its species, on the floor of the MSE Library,” or however more specific one would care to make it.

There is at another way in which imagination_P enjoys an advantage over imagination_O. Were some person *A* able to imagine_O some *P*, it would be somewhat difficult to accurately (or at least fairly accurately) capture all relevant details of the imagined *P* so that *A* could use this imaginability as evidence of *P*'s possibility to person *B*. For instance, suppose I am trying to explain to a friend of mine who has never read *The Lord of the Rings* the appearance and general features of the hobbits. I might say that they have hairy feet, rosy cheeks, plump bellies, and soft hands. I might be able to give a very accurate description of my picture of a hobbit, but this is no guarantee that, upon my description, my friend will have the same picture of what a hobbit looks like that I do in the quasi-sensory sense.²⁰ But if we're interested in imagining_P, in this case,

²⁰A very vivid example of this can be seen if one compares the old animated Rankin-Rass portrayal of hobbits with those given in Peter Jackson's live-action trilogy.

this should be good enough. If I describe to him my $image_P$ of a hobbit, it's unlikely he will get a relevantly (if at all) different $image_P$ of a hobbit. One seems to stand a much better chance of giving a justification of one's beliefs about conceivability, then, if one uses $imaginability_P$.

2.2. The rationalist account. In contrast to the account presented in the previous paragraphs, the rationalist account is less interested in the ways the agent might employ his imagination. Instead, if a rationalist modal epistemology cares at all about conceivability, the kind it wishes to examine will involve the examination of a particular proposition or expression, possibly in conjunction with one or more other propositions or expressions, with an eye to determining their (nomological, metaphysical) compatibility. Additionally, where the empiricist account is, at least on some picture, in principle concerned with some sensory or quasi-sensory activities (for instance, $imagining_O$), the rationalist account is primarily concerned with what can be discovered a priori. Pretty naturally, then, any sensory or quasi-sensory component is of tangential concern, at best. What matters in this case is the correct application concepts or principles which are (in some sense) a priori.

One shouldn't get the impression that rationalist accounts of modal knowledge are totally homogeneous. For instance, Peacocke (1999) argues that correctly understanding modal logic—and consequently achieving modal knowledge—involves the correct application of what he calls the “Principles of Possibility.” Under this

account, “a specification represents a genuine possibility only if it respects these principles.”²¹ He continues:

My claim is that someone who understands modal discourse has a form of implicit knowledge of the Principles of Possibility. It is this implicit knowledge that allows him to discriminate between those ways which are ways the world might be, and those ways or specifications which are not. More generally, someone who understands modal discourse applies this knowledge of the Principles of Possibility in evaluating modal claims.²²

Chalmers (2002), on the other hand, offers a broadly rationalist account which fits nicely into our current discussion:

Negative notions of conceivability hold that S is conceivable when S is not *ruled out*. For example, a sense of “conceivable” in common usage holds roughly that S is conceivable when it is not ruled out by what one knows, or by what one believes [...] The central sort of negative conceivability holds that S is negatively conceivable when S is not ruled out a priori, or when there is no (apparent) contradiction in S[...] We can say that S is *prima facie* negatively conceivable for a subject when that subject, after consideration, cannot rule out S on a priori grounds. And we can say that S is ideally negatively conceivable when it is not a priori that $\sim S$.²³

The chief takeaway is that there is some variety of opinion about just what the rationalistic project in modal epistemology should entail. As we’re concerned with views which think conceiving of

²¹Peacocke (1998), p. 125.

²²ibid., 125–126.

²³Chalmers (2002), p. 159. Emphasis mine. This is only one amongst many notions of conceivability that Chalmers discusses; indeed, positive conceivability is a sort of empiricist account, and the negative/positive connection is discussed at length.

something is in some way constitutive of knowing that it is possible, we'll only be examining Chalmers' view here.²⁴

The description of ideal negative conceivability seems pretty appropriate for our current discussion. We might collect this notion into the following definition:

DEFINITION 9. (Negative conceivability) *A can conceive_n of P iff A cannot rule out P on a priori grounds.*

Here a priori grounds means what it typically does—grounds which make no reference to methods or concepts exclusively reliant on sense data.

Here, as before, we run into the question of subjectivity. The description makes reference to whether or not P is conceivable for a particular subject if such-and-so constraints hold and certain conditions are met. This may give us a good idea of whether any *particular* judgment is true. What it does *not* do is furnish us with a general principle—and that's the project of this section. Furthermore, we run into the problem of whether or not the agent of question is capable of understanding the implications of all (true) beliefs he has, or whether he'll be able to effectively carry out such a process of comparison or inspection.

This suggests a natural formulation of the strong version of NC, along the same lines that we did previously with the "ideal conceiver".

²⁴Arguably, we could try and coax a conceivability principle from Peacocke's view. But that would be eisegetical, so we won't do that.

DEFINITION 10. (Ideal Negative Conceivability) *P* is ideally negatively conceivable (or “conceivable_{in}”) iff *I* cannot rule out *P* on a priori grounds.

The question of what precisely comprises the ideal conceiver’s basis set of beliefs is still relevant. One might think that the precise content—or more correctly, the size—of the belief set doesn’t matter very much. After all, the relevant question is compatibility, not inferential strength (that is, the number of true beliefs the agent can acquire by examining his priori commitments). I think this is a mistake. Take, again, our hypothetical 6th grader. Suppose that we endow him with a belief set containing all theorems of number theory, set theory, model theory, algebra, and the rest of mathematics, couples with knowledge of all proof strategies and rules of inference. Even so equipped, our hypothetical 6th grader will not be able to rule out, on a priori grounds, that the nominative plural case of the first declension Latin noun *nauta* is *nautarum* (in fact, it is *naute*).²⁵

Granted, this is not an instance of making an incorrect modal judgment, but rather an incorrect judgment about Latin grammar rigidified to the world which the 6th grader inhabits. But the claim was that conceivability_{in} will render certain things conceivable which are not, in fact, true, provided that the ideal conceiver is only equipped with a non-maximal basis set of beliefs. Consider a modally stronger example. If we extend our 6th grader’s knowledge

²⁵Operating under the assumption, of course, that our 6th grader is not well-versed in Latin; if he were, the case would be trivial, but we could move over to Attic Greek and be in an equally good case.

base to some physics—say, physics up until the year 1895—he will probably have at least one modally false belief: a universe can exist which obeys Maxwell’s equations under Galilean transformations, rather than Lorentz transformations, in having a maximum possible velocity of c .²⁶ Because the theory of lumeniferous aether was demonstrated to be incorrect *empirically* by the Michelson-Morley experiment (and empirical experimentation is the paradigm case of an a posteriori procedure), it is unlikely that only an a priori analysis of the pre-existing body of physical knowledge is enough to render it false.

The upshot of this is the following: the basic belief set of \mathcal{I} is going to have to be expanded. In particular, it will have to be expanded to a sufficient level where no false beliefs will be coherent with it.²⁷ That there even could be such a basis is a controversial position, since accepting this position comes very close to accepting a fully rationalistic account of knowledge. For the moment, we will assume that such a thesis is true. If the reader finds this assumption problematical, then he can follow our discussion in a spirit of thought-experiment: *if* these assumptions and theses are true, then such-and-so results follow.

²⁶Giving a full explanation of this point would take us too far afield. If the reader wishes to read a full exposition of the argument, he may find them in a standard textbook on special relativity, say [Resnick \(1968, pp. 16-18\)](#).

²⁷One might want to add a “compactness” condition, and say that no false beliefs will be coherent with any subset of his basic belief set. We won’t explore this possibility here.

Again following Chalmers (who in turn follows the *Aufbau*), we can cash out the desired basis belief set in a way similar to what he calls a scrutability base:

A scrutability base is a class of truths from which all truths are scrutable, for a given notion of scrutability[. . .]How small can a scrutability base be? Let us say that a minimal scrutability base is a class of sentences C such that C is a scrutability base and no proper subclass of C is a scrutability base.²⁸

Scrutability, generally speaking, is the thesis that, given a certain class of truths, one can discover—through inference or some other process—all other truths of the world.²⁹ In our case, as well, this process of discovery is purely a priori. What precisely that class of truths happens to be is our current concern. The above description is somewhat more extravagant than what we need for the present purposes. At the moment, we are concerned with what sorts of true beliefs one would need in order to determine truths about what is possible. For instance, it seems logically possible that the fine structure constant should have a value of (approximately) $\frac{1}{136}$ rather than (approximately) $\frac{1}{137}$, but the class of truths from which it is conceivable_{in} that such a state of affairs is possible differs substantially from the class of truths from which it is conceivable_{in} that state of affairs is actual—especially since the value of the fine structure constant in the actual world is in fact (approximately) $\frac{1}{137}$.³⁰

²⁸Chalmers (2012), pp. 38–39.

²⁹Chalmers has (at least) 4 different types of scrutability: inferential, conditional, a priori, and generalised. For more detail, see the Introduction and §2.2-2.6 of Chalmers (2012).

³⁰Its exact value is $\alpha = \frac{k_e e^2}{c\hbar} \approx 7.2973526 \times 10^{-3}$.

Consequently, we might be well off in adding another detail to our previous description of the ideal conceiver:

DEFINITION 11. (Epistemic Soundness and Maximality) \mathcal{I} possesses a relevantly maximal set of true beliefs.

The notion of a relevantly maximal set of true beliefs is similar to that of a scrutability base. We can try to give it a more precise definition: let S be a scrutability base from which all categories of truths (microphysical, modal, logical, and so on) are scrutable. Then a relevantly maximal scrutability base is a subclass $M \subset S$ from which all truths of a given category are scrutable. If the scrutability base is a minimal scrutability base, then M is unique.³¹ A relevantly maximal set of true beliefs, then, is a set of beliefs whose propositional content corresponds to the members of M .

3. Recap

A brief refresher on the main points of this chapter is in order.

- (1) If we want to use RR to establish more solid, credible results on modality, it is best to construct it as an entailment relation—conceivability entails metaphysical possibility—rather than just a reliability relation—conceivability offers a better-than-not guide to metaphysical possibility. We can render this entailment as $Ca\phi \rightarrow \diamond\phi$, where $Ca\phi =_{af} Ja\diamond\phi$, or in English, “ ϕ is conceivable if a judges ϕ to be possible,” for a suitable notion of “judges.”

³¹This may need to be proven a little further, but it has a certain degree of prima facie appeal.

- (2) If we want to have the strong version of RR, then it is best to construe the conceivability part of the entailment in terms of an ideal conceiver. This ideal conceiver plausibly has the following characteristics:
- (a) He possesses ideal mental and cognitive capacities. (Ideality)
 - (b) He possesses a relevantly maximal set of true beliefs. (Epistemic Soundness and Maximality)
 - (c) He is capable of completing any rational or extrapolative process from his basic belief set in a finite number of steps. (Finiteness)
 - (d) The set of all propositions consistent with his basic belief set is semi-decidable. (Checkability)
- (3) There are roughly two categories into which accounts of conceivability can fall.
- (a) *Empiricist*. These accounts of conceivability typically hold that conceivability involves some sort of mental exercise which has essential *sensory or quasi-sensory* qualities. This imagination can be either *propositional*—to use Yablo’s example, imagining that there is a tiger behind the curtain—or *objectual*—following Yablo again, imagining the tiger itself.
 - (b) *Rationalist*. These accounts of conceivability typically do not require that conceivability have any *essential* sensory or quasi-sensory qualities. Instead, these accounts tend to revolve around an agent examining

some proposition to see whether it passes certain sorts of tests; typically, *contradiction-detection* is the favorite.

CHAPTER 2

Antinomies of Pure Conception

The second class of sophistical arguments is occupied with the transcendental conception of the absolute totality of the series of conditions for a given phenomenon, and I conclude, from the fact that I have always a self-contradictory conception of the unconditioned synthetical unity of the series upon one side, the truth of the opposite unity, of which I have nevertheless no conception. The condition of reason in these dialectical arguments, I shall term the antinomy of pure reason.

– Immanuel Kant, *Critique of Pure Reason*, A 340/B 398

In the previous chapter, we examined various forms of the conceivability-to-possibility arguments, for differing values of “conceivable.” It’s possible, however, to make the argument stronger, either by *identifying* possibility with conceivability, or else to add to RR the converse—that is, $\Diamond\phi \rightarrow C\phi$ again for the specified value of “conceivable.” We can call—and will henceforth call—theses of this kind “modal monism,”—following [Vaidya \(2008\)](#)—since they hold (with some qualification) that the space of metaphysically possible worlds and epistemically possible worlds are coextensive. Additionally, we will mean by “modal monism” the purely conceivability-based version.

The argument of this chapter is, in short, that modal monism is false, or at least that multiple versions of it are false and others are quite debatable. The converse of RR leads us to conclude that certain

things, which we have decent independent reasons for thinking possible, are in fact inconceivable and therefore impossible. The problem, it appears to me, is that if we *identify* the possible with the conceivable—which is an important part of this version of modal monism—then the two must have the same range—that is, everything which is possible is also conceivable, and contrariwise. As will be argued in this chapter, this is an implausible—or at least quite fraught—proposition.

The first order of business, then, is to get a good hold on the different types of monism. This is the topic of the first section.

0. Modal monism

Building on the work of Chapter 1, we might construct our desired version of monism by adding an another clause to RR. There are two broad ways we could construe this clause. These are, roughly speaking:

- (1) The implication thesis: $\Diamond\phi \rightarrow C\phi$
- (2) The identification thesis: $\Diamond\phi \equiv C\phi$

0.1. Expansions of (1). Does possibility entail conceivability? At first blush, one hard pressed to know how to answer. There are at least two reasons for this. First, it's not precisely clear just what sense of possibility we are concerned with: do we care whether a given thing is *logically* possible? *Metaphysically* possible? Second, it's likewise not clear what notion of conceivability we are concerned with. As a result, it's difficult to know whether a blanket answer to that question is possible. We will consider both points in turn.

First, what notion of possibility are we interested in? Broadly, there are three: logical, metaphysical, and nomic. Logical possibility consists in not implying a contradiction or being logically inconsistent. Metaphysical possibility, while perhaps a little less expansive than logical possibility, is still quite broad. In order to be metaphysically possible, an entity must not only not contain (or imply) any logical contradiction, but must also conform to any broad metaphysical principles which may hold across all worlds. For instance, some essentialists hold that a given property *P* of a substance *S* is an essential property if there is no situation in which *S* exists where *P* is not exemplified. Consequently, it is metaphysically—though not necessarily logically, depending on the analysis one wishes to conduct—impossible for there to be an instantiation of *S* without a concurrent instantiation of *P*. Finally, we have nomic possibilities, which correspond to those varieties of modality connected with *laws of nature*. To turn to an earlier example, it would be a question of nomic possibility whether or not there could be a universe which is accurately described by Maxwell’s equations under Galilean transformations. Similarly, whether there could be a universe which is totally accurately described by both modified Newtonian dynamics and physical theories involving dark matter would be a question of nomic possibility.

In principle, it would be possible to construct a version of (1) for any of these three. For the time being, we will concern ourselves with logical possibility. To a first approximation, we can say the following:

(1) If P is logically possible, then P is conceivable.

If we combine this question with the different kinds of conceivability discussed so far we arrive at:

(1.1) If P is logically possible, then P is conceivable_O.

(1.2) If P is logically possible, then P is conceivable_P.

(1.3) If P is logically possible, then P is conceivable_{in}.

Note that the more natural choice for this would be (1.3) (or possibly (1.2)), since all it involves is contradiction detection. Determining broadly *logical* consistency from some sensory or quasi-sensory mental representation seems, failing some immediate failure (along the lines of a square circle or a sphere both red and green all over), a difficult task.

0.2. Expansions of (2). (2), in its strongest form, identifies the possible with the conceivable. How ought we construe “conceivable” here? Following the categorizations in the previous chapter, we have roughly three options: we construe it as an act of imagination_O, an act of imagination_P, or as an act of conceiving_{in}. We will take the possibility in question to be *logical* possibility, as in the previous section. This would leave us with the following principles:

(2.1) The possible is the conceivable_O.

(2.2) The possible is the conceivable_P.

(2.3) The possible is the inconceivable_{in}.

Suppose we think that the possible is to be identified with the conceivable_O. The relevant principle would then be (2.1). The fact

that a tiger with 79 black stripes and a lame left leg is possible, say, is then the same as the fact that a tiger with 79 black stripes and a lame left leg is conceivable_O. But just who is doing the imagining_O? The natural choice is our ideal conceiver, *I*.

There's an immediate question regarding considerations of temporality. For a significant period of time, there were no human minds in the universe—the better part of the life of the universe, in fact. One might suppose that there are other minds in the universe besides humans, or at least that there *were* such minds for some period of time. In any event, there was very plausibly a time where there existed no conscious minds of the kind we are familiar with at any place in the universe. If we identify the possible with the conceivable_O, and this last makes reference to the idea of an ideal conceiver as something *other* than a mere heuristic, we might think that, prior to the advent of conscious minds of a type sufficiently similar to the ideal conceiver, there were no entities, states of affairs, or propositions that we might call “possible.” And a position which holds that the existence of such conscious entities at one time or another might retroactively determine possibilities in the early universe, or in some other, non-actual world is—to put it charitably—a strange one.

The question of specificity, discussed earlier, is relevant to the current question. Recall: imaginability_O was not in a good position to make a judgment concerning matters of specific, fine-grained properties of the imagines state of affairs. Consequently, should we identify possibility with this kind of conceivability, conceiving_O of

such a situation might not eliminate the possibility of some unanalyzed contradictory or inconsistent property or concept. Just how much specificity would be required is unclear. But this is essentially a procedural worry, not a substantive one. Since our concern is with the connection between the possible and the conceivable, we won't pay too much attention to this worry.

If we identify the possible with the conceivable_P, the relevant principle would be (2.2). This version enjoys the same advantage over (2.1) that conceivability_P does over conceivability_O. Since the ideal conceiver would be imagining, in some sense, a propositional description of the 100 lb. trout (say), specificity would be a non-issue. A constructive procedure for this activity might run as follows. One starts with the regular conception_P of a trout of the relevant kind. One then specifies the exact changes one would have to make to the original conception in order to get at the desired *possibilium*, all the while ensuring that the changes made do not produce any conceptual inconsistencies. The procedure would be similar for any potential *possibilium*, though the degree of complexity vary substantially.

1. Against modal monism

Up until now no argument against modal monism has been offered. That endeavor is the task of this section. First, we will discuss some pre-existing arguments in the literature against certain directions of the argument. Subsequently, we will present additional arguments for this position.

1.1. A posteriori necessities. One class of counterexamples are based on so-called a posteriori necessities. The usual examples involve some conceptual truth that is supposed to be discovered a posteriori, such as “Hesperus is Phosphorus” and “water is H₂O.” These sorts of arguments can be schematised as follows:

- (1) S is possible iff S is conceivable
 - (2) P is necessary and a posteriori
 - (3) Σ is a situation in which it is the case that $\neg P$
 - (4) Σ is conceivable
 - (5) Σ is possible
 - (6) $\neg P$ is possible
 - (7) P is not necessary
 - (8) \perp
-

\therefore It is not the case that S is possible iff S is conceivable

For instance, it is commonly held that it is a conceptual necessity that water is H₂O. Yet it is also sometimes accepted that we can conceive of scenarios which would falsify one’s belief of water’s being H₂O. In a famous example due to [Putnam \(1975\)](#), we are asked to imagine a “Twin Earth.” The inhabitants of this Twin Earth refer to a certain chemical compound as water. It does everything water does. But it’s chemical formula is *not* H₂O, but rather “XYZ.” We can, or so the claim goes, “conceive” of this situation a priori. But it is in fact an impossibility that water should *fail* to be H₂O, we conclude that conceivability is no good guide to possibility—or so at

least the argument goes. This argument targets RR specifically; it leaves unattended whether possibility entails conceivability.

Further, this is a counterexample directed against a particular version of conceivability, one in which the agent in question conceives of a situation which he takes to verify a certain proposition. This differs slightly from some of the applicable versions of conceivability which we entertained above (conceivability_O). In this situation, the Twin Earth example is a scenario which is possible *for all you know*, an which is taken to falsify some proposition P. Conceivability_O, on the other hand, invites the agent to *actively imagine* the scenario in question. Yablo puts it this way:

p is conceivable_{ijb} iff one can imagine justifiable believing that *p*[...]What I have been calling *philosophical conceivability* is none of these.¹

What he calls *philosophical* conceivability divides up into conceivability_P and \neg _O.

What are we to make of the counterexample in light of the distinction? Perhaps imagining_O that water is not H₂O will involve some sort of contradiction in concepts, and thus won't leave itself open to this counterexample.² By happy consequence, one might maintain, when one attempts to imagine_P the given scenario by starting from the simple case—something being water or something being H₂O—one simply cannot arrive at the desired counterexample.

¹Yablo (1993), p. 26. Emphasis mine.

²Note that conceivability_P leaves itself open to as much propositional depth and specificity as one happens to need at the moment.

Similarly, Chalmers argues that such counterexamples do not count against his notion of conceivability (quite close to what we have denoted conceivability_{ip}):

The other standard source of gaps between conceivability and possibility arises from Kripkean cases. It is often said that it is conceivable that Hesperus is not Phosphorus, or that water is not H₂O, or that heat is not the motion of molecules, but none of these states of affairs is in fact possible. In these cases we have a posteriori necessities and impossibilities, out of the reach of a priori methods.

There are a couple of things to be said here. Clearly, the main sense in which these states of affairs are conceivable involves primary conceivability. As discussed earlier, the states of affairs in question are not secondarily conceivable. At best, they might be *prima facie* secondarily conceivable for a subject lacking relevant empirical knowledge. They will not be *prima facie* secondarily conceivable for a subject with the relevant knowledge, and they will not be ideally secondarily conceivable as that notion is spelled out above.³

Here he employs his distinction between primary and secondary conceivability made a section earlier:

We can say that S is *primarily conceivable* (or *epistemically conceivable*) when it is conceivable that S is *actually* the case. We can say that S is *secondarily conceivable* (or *subjunctively conceivable*) when S conceivably *might have been* the case. This corresponds to two different ways of thinking about hypothetical possibilities: epistemically, as ways the world might actually be, and subjunctively, as counterfactual ways the world might have been.⁴

³Chalmers (2002), p.161.

⁴ibid., p. 157.

Making use of this distinction, one can argue that the state of affairs corresponding to the description “H₂O is not water” is not secondarily conceivable_{in}. On a priori grounds, an ideal conceiver as defined in the previous chapter would be able to rule out since this is a conceptual inconsistency, akin to the description “humans are not mammals.”

Another doubt one might have is something like the following: the notion of conceivability expressed by secondary conceivability_{in} is by nature a priori; the possibility or impossibility of these sorts of counterexamples is itself a posteriori, and thus not determinable on grounds independent of experience. Consequently, the possibility or impossibility of these cases cannot be determined on solely a priori grounds.

This is partially correct and partially mistaken. Certainly, the modal status of such counterexamples is determined—or at any rate *was* determined—by a posteriori means. But this is only a problem if the base belief set of our ideal conceiver is filled only with those truths whose factive status may be determined a priori, and there seems no reason to suppose that this is the case.

In this sense, the *methods* of discovery or deduction are quite close to being a priori, but the data from which the methods proceed may be discovered or acquired through whichever means happens to be available. Since the ideal conceiver’s belief base will be maximally complete, one might expect that his concepts of “water” and “H₂O” will themselves be maximally complete (in the sense of maximally well-specified). As a consequence, he would be able to rule out

using only a priori methods the possibility of a scenario which is accurately describes by “water is not H₂O.” This, in turn, is what is required to determine whether or not the given scenario is conceivable. Since he can rule it out a priori, he (and we) might conclude that it is not.

Certainly, this is a somewhat modified notion of the a priori (call it “a priori +”). But I take it to be relevantly similar to the classic notion in the following way: while the concepts which \mathcal{I} is employing may in principle be supplied by the senses, all the conclusions reached are reached by means of conceptual analysis and comparison, both of which are paradigmatic aprioristic endeavors. The fact that gold has atomic number 79 is a truth which is determined empirically (that is, a posteriori). But it is no less a conceptual truth for all this. As a result, in idealized circumstances, a reasoner in possession of the complete concept of “gold” will (or so is the claim) be able to determine the truth of the statement “gold has atomic number 79”.

We can also consider the converse, $\diamond\phi \rightarrow C_ax$. Clearly, P is not possible, as it is necessarily true. It would be a mistake, however, to infer from this that P is not *conceivable*; this would be to affirm the antecedent. Consequently, somebody who accepted (1) but not RR could maintain that these sorts of counterexamples are in fact conceivable though not possible and not incur any inconsistency in his ideology.

So while the a posteriori counterexamples do in fact apply to certain types of conceivability, the ones which we have articulated

in this section are immune from them. We now turn to other classes of counterexamples.

2. *Inconceivable!*

For the sake of completeness, we'll lay out the following—weak—version of this hypothesis:

DEFINITION 12. ($\neg 1$) *If P is inconceivable for some ideal conceiver \mathcal{I} , then P is impossible.*

Again, this can be cashed out in terms of the different sorts of conceivability discussed until this point:

($\neg 1.1$) If P is inconceivable_O for some ideal conceiver \mathcal{I} , then P is impossible.

($\neg 1.2$) If P is inconceivable_P for some ideal conceiver \mathcal{I} , then P is impossible.

($\neg 1.3$) If P is inconceivable_{in} for some ideal conceiver \mathcal{I} , then P is impossible.

We will take each in turn.

2.1. Does Inconceivability_P Entail Impossibility? ($\neg 1.1$) holds that if an ideal conceiver, as construed above, finds P inconceivable_P, then P is impossible. What would this mean? Remember that \mathcal{I} finds P conceivable_P when \mathcal{I} has a propositional attitude whose content yields an accurate description of P. Correspondingly, \mathcal{I} would find P inconceivable_P if \mathcal{I} cannot have a propositional attitude whose content yields an accurate description of P.

To see why, let's specify the following scenario:

Let I be an ideal conceiver. Let w_f be a world in which the Goldbach conjecture is false. Then w_f is possible iff \mathcal{I} can conceive_P of w_f . As a corollary, if w_f is inconceivable_P for \mathcal{I} , w_f is impossible.

Note that a well-stated theorem is either necessarily true or not at all. So if \mathcal{I} can, in fact, conceive of w_f , then w_f is possible, and consequently the Goldbach conjecture is (necessarily) false. On the other hand, if w_f is inconceivable_P for \mathcal{I} , then w_f is impossible. Well, so what? It might be. The problem is that, given certain descriptions, we can easily stipulate the following scenario as well:

Let \mathcal{I} be an ideal conceiver. Let w_t be a world in which the Goldbach conjecture is true. Then w_t is possible iff \mathcal{I} can conceive of w_t . As a corollary, if w_t is inconceivable_P for \mathcal{I} , w_t is impossible.

Now, all it really takes to conceive_P of either w_f or w_t is to imagine the following: " w_f/w_t is a world exactly like ours, except that the Goldbach conjecture is false/true". And conceiving_P of this is child's play. We don't even need \mathcal{I} to do it. Which means that, if we're fine with using both types of conceivability in our definition, we can very easily conceive_P of w_t and w_f simultaneously. Which is to say, \mathcal{I} can conceive_P of the Goldbach conjecture being simultaneously necessarily true and necessarily false.

One might respond that, since \mathcal{I} has a base belief set that contains only *true* beliefs, he cannot in fact conceive_P of a world in which a contradiction holds. Consequently, the situation described above is not coherent, which entails that the principle still holds.

I don't think this response succeeds. All that it takes for \mathcal{I} to find P conceivable_P is for \mathcal{I} to be able to have a propositional

attitude which yields an accurate description of P. And, while a world containing a contradiction is not a *possible* world, it is in fact a *conceivable_p* world in this sense.

It gets worse. We can, in fact, have something like the following:

Let \mathcal{I} be an ideal conceiver. Let w_{\perp} be a world in which the Goldbach conjecture is both true and false. Then w_{\perp} is possible iff \mathcal{I} can conceive_p of w_{\perp} .

Which is to say: \mathcal{I} can conceive_p of an impossible world. This does not strictly count against (\neg 1.1), but against the general version of RR. The argument is that, if we use conceivability_p, then RR overgenerates possibility claims. In fact, it generates possibility claims that are, by most lights, necessarily false.⁵

One could, if one wished, take this even further. The sentence “the world at which the sentence ‘every even number greater than 2 can be written as the sum of two odd primes’ is either true or false is inconceivable_p” (call this S) seems to yield some *intelligible* propositional content, at least to the extent that it makes use of concepts which, we’re able to see, are contradictory. So, we might suppose, \mathcal{I} might be able to conceive_p of a world at which it’s true. But that lands us in the unenviable position of being committed to the position that \mathcal{I} can conceive_p of a world where S is true and where the sentence “it is true that every even number greater than 2 can be written as the sum of two odd primes” is also true. Which means that \mathcal{I} is again conceiving_p of an impossible world.

⁵This argument can be lodged against certain other modal epistemologies. For a similar argument against some theses of Timothy Williamson, see [Roca-Royes \(2011b\)](#). The phrase “overgenerates possibility claims” is, in fact, hers.

One might be tempted to claim that, because this is in fact an impossible world, \mathcal{I} would not be able to conceive_p of it. After all, \mathcal{I} is supposed to have only true beliefs in their base belief set, and presumably this would include the truth or falsity of the Goldbach conjecture. If \mathcal{I} were to conceive_p of w_{\perp} , he will in fact be conceiving_p of a contradiction, which seems incompatible with the notion of him possessing a base belief set containing only true beliefs.

There is both an easy and a hard response to this argument. The easy one is that this is just the re-assertion of the relevant version of RR (or, more accurately, of one of its consequences, $\neg\Diamond\phi \rightarrow \neg C_a\phi$), since the argument can't get off the ground without the assumption that impossible things are not conceivable_p. Consequently the response begs the question. But maybe this is overly flippant

The hard response is that we *seem* to be able to conceive_p of what are—under regular assumptions—“impossible” situations. For instance, some philosophers have argued that (a very restricted class or realm of) contradictions are rationally believable.⁶ The suggestion is not that they are correct. It is that the fact that there is a very detailed account of just *how* contradictions might under certain circumstances be true counts as evidence that we in fact can conceive_p of “impossible worlds.” And the fact that we can *understand* what such people say and make sense of their arguments—perhaps even while dismissing it as “monstrous moonshine”—counts as some evidence that these arguments are at least *intelligible*, even if false.

⁶For an example see [Priest \(1998\)](#).

2.2. Does Inconceivability_O Entail Impossibility? (\neg 1.2) holds that if an ideal conceiver \mathcal{I} finds P inconceivable_O, then P is impossible. To expand a little, if \mathcal{I} cannot have a sensory or quasi-sensory mental representation (or suitable analog of a mental representation) of P. Naturally, \mathcal{I} has all the properties attributed to him in the past chapter.

Since the converse of RR is supposed to be a sort of entailment relation as well, if we want to find a counterexample, we need to find a case in which the antecedent is true and the consequent is false. In other words, we need to find a case where something is (metaphysically) possibly instantiated but is simultaneously inconceivable_O.

Here's a first shot: human vision only takes in a very, very small part of the electromagnetic spectrum (wavelengths between 390-700 nm). Anything on either the UV or IR ends of the spectrum are consequently not visible. However, it is—or so runs the contention—metaphysically possible that a human might have a visual system that would be able to see other parts of the spectrum—say, with a lower-end expansion of 200 nm or so. Further, the contention is that this other human will have a series of accompanying phenomenal experiences totally closed off to humans—like the author and the reader⁷—who have the visual capacities common to humans.

This limitation extends to our ideal conceiver. We argued earlier that the ideal conceiver should only be endowed with—maximally excellent—human faculties, since we are interested in what modal

⁷Presumably.

knowledge *we* might come to have through some sort of conceivability process. Consequently, \mathcal{I} will have only the human capacities for phenomenal experience. Since \mathcal{I} has no access to the phenomenal experience of seeing “ultra-yellow” or “infra-blue” or some other appropriately-named new color, he will not be able to have a sensory or quasi-sensory mental representation of these colors, which is the same as saying that he finds that (phenomenal) state of affairs inconceivable_O.

This argument can be pretty easily reproduced for different sorts of inconceivable_O situations; all that would be required for the new argument would be to represent a different inconceivable_O by P. The natural question, then, is whether or not there are other instances of this kind on offer.

Another possible argument of this type comes to us from quantum mechanics. Suppose that we have a fermion of spin- $\frac{1}{2}$ with state vector:

$$|\psi\rangle = \frac{-i}{2} |+\mathbf{z}\rangle + \sqrt{\frac{3}{4}} |-\mathbf{z}\rangle$$

where $|\pm\mathbf{z}\rangle$ are basis vectors of a Hilbert space. It’s possible to get a decent intuitive hold on what “spin” is (as many physics undergraduates discover during their second or third year of study), even though there is no classical analog to this phenomenon. It may even be possible to have an accurate phenomenal representation of what fermion spin is like.⁸ For the sake of argument, let’s suppose that \mathcal{I} might have an accurate sensory or quasi-sensory mental

⁸For instance, I have a mental picture of what spin *might* be like, but I don’t think it’s terribly accurate, since for most purposes fermions have all the extension of a geometrical point.

representation of what spin is. Even if this is so, however, it is hard to see just how \mathcal{I} might be able to make sense of the fact that $|\psi\rangle$ has, not a *definite* spin, but instead a superposition of the different spins a single fermion of the relevant spin might have.

By analogy, it's difficult to get a good intuitive grasp of torque, at least at first. After a bit of work, however, one can develop the requisite intuition. It is hard to make sense of how one might develop the intuition of a rotor, for instance, experiencing a superposition of torque. This is not a matter of simple mental laziness or lack of cognitive success, but instead a matter of not having the correct concepts or mental capacities for being able to get a picture of this. In the case of torque, it is not terribly difficult to see why. The torque experienced by an object is given by the equation $\tau = \mathbf{r} \times \mathbf{F}$, where \mathbf{r} is the vector which gives the position of the object under consideration, and \mathbf{F} is the force the object experiences.

As a measurable, physical quantity, produced by other measurable physical properties, it makes little *intuitive* sense to talk about a superposition of torque, as if an object were undergoing a torque of 50% 60 N·m and 50% 75 N·m. Since we ordinarily brush up against the macroscopic world, which works (to a very good local approximation), we might expect it to shape our pre-philosophical intuitions and our concepts, to a large extent. Against that background, it is unsurprising that we should find states of affairs *not* describable in classical terms inconceivable_O.

One might object—not unreasonably, perhaps—that basing an argument about what our powers of conceiving_O can and can't do on

so esoteric a notion as spin does a bit of violence to the intended goal. Maybe this is true, and maybe not. But for the sake of argument, let us grant that there is some merit to this counterargument. Is the defender of conceivability_O out of the woods yet?

No. Take a concept which is a good deal less abstract and apprehensible: spatial position. It turns out that we can run a very similar argument with this description of a quantum state. Consider our old friend the fermion of spin- $\frac{1}{2}$. The general form of $|\psi\rangle$'s representation in position space is

$$|\psi\rangle = \sum_{i=0}^N |x_i\rangle \langle x_i|\psi\rangle$$

where N represents the number of possible positions the particle could occupy, and the $|x_i\rangle$ are basis vectors of the position space. In much the same way as before, the particle is thought to in some sense be in all the possible spatial configurations of the system at once. From this we make much the same argument as before, only with what seems like more force. It *might* be possible, once one develops one's concept of spin more rigorously, that one might be able to have the relevant phenomenal experience that would count as conceiving_O. But in the case of spatial position, this option doesn't seem to actually be an option: we actually *have* the concept of spatial position, and it doesn't seem to admit of bi-location.

In the last two examples, there was a tacit assumption: namely, that some version of scientific realism is true. Under this view, the elements of our best physical theories do indeed describe the real world; not only our phenomenology (in the physical sense), but

also our mathematical formalisms are thought to have thoroughly metaphysical implications—though naturally there can be a range of disagreement as to what *precisely* those implications are. So, in a very real sense, this argument will stand or fall according to whether or not we think that our physical theories do carry those sorts of implications. Consequently, if one is a scientific realist, one should accept this argument or give up scientific realism. The scientific anti-realist, however, need not accept this version of the argument.

There are a few ways one could counter this argument. First, one could say that the state of affairs described by P is not possible. In the two cases offered above, this can be accomplished by, first, denying that humans slightly differently constituted could possibly have the relevant phenomenal experiences; and second, by either denying that the states of affairs described are in fact inconceivable_O or by adopting a certain sort of scientific anti-realism.

The first option isn't a very promising one. It makes the particular sort of vision that humans have a(n) (modally) essential property of *Homo sapiens*, which does not seem like a very promising line of argument. Certain things seem like they might count as essential properties of a human—being a featherless biped, for instance—but a very specified, particular sort of vision doesn't seem to be one of them. The first part of the second option seems equally unpromising. Moreover, the concept of spatial position works against him in this instance. Continuous objects may in some sense be said to “be” in multiple spatially connected but distinct places at once, but such is not (to the best of my knowledge) the case with discrete objects.

As to the second part of the second option, this will, as indicated above, depend upon whether he is a realist or not.

Another option is to deny that \mathcal{I} really finds P inconceivable_O. This seems like a weak line of attack, mainly because of how we construed the ideal conceiver in the first place. Speculative philosophy is only as good as the methods we can use to arrive at its theses; if such a method is unusable, or is only useful in very particular circumstances, the theses based upon it fall. If one were to deny that \mathcal{I} finds P inconceivable_O, one would need to impute him faculties that, to the best of my knowledge, no human possesses. It's reasonable to believe that such entities possibly exist, but basing this notion of conceivability_O on just such an exemplary entity does little to aid the modal metaphysician or epistemologist. At best, this action would render conceivability_O arguments an appeal to (hypothetical) testimony, and this seems a tenuous ground of knowledge at best.

Behind the preceding discussion of "quasi-sensory" states and so on hangs the question of what these states comprise. In particular, which concepts does \mathcal{I} employ in his exercises?⁹ This is a deep question, but not one which is the topic of this thesis, and so its treatment will be brief. The key question is just how we get our

⁹See [Menzies \(1998\)](#) and [Jarvis \(2012\)](#) for this sort of concept-based theory of imagination. The former ties modality as such into conceivability. Ichikawa and Jarvis mostly treat with a notion of conceptual possibility. In their language, the argument of this section amounts to a denial of MMR from right to left. Peacocke ([1997](#), [1998](#), [2002](#)) defends a principle-based account of modality which is not based on conceivability (at least not explicitly). Here the important thing is not the concepts being employed, but the rules (the "Principles of Possibility") according to which they are assessed.

concepts. If one thinks that these concepts are derived, at the last, from sensory experience, then our acts of conceiving_O (and therefore *I*'s acts of conceiving_O) bottom out on objects acquired via our sensory field. And we are not the sorts of creatures which have sensory apparatus keen enough to form concepts by experience of the sub-atomic world. This brings accounts of this sort under fire from the discussions above.

2.3. Does Ideal Negative Inconceivability entail Impossibility? Ideal negative conceivability holds that *P* is conceivable_{in} by *I* just in case *I* cannot rule out *P* a priori. Another way of saying this is that $\neg P$ is not a priori. On the view of people who defend this or a related version of conceivability, there is an important relation between the necessary and the a priori. For instance, Chalmers argues that while perhaps not all modal truths are a priori (recalling, for instance, Kripke's arguments to the contrary), all modal truths are at least scrutable from truths determinable a priori:

[W]hat about metaphysical modality? Here it is familiar that truths such as 'it is necessary that water is H₂O' and 'It is necessary that Hesperus is Phosphorus' can be true without being a priori. In these cases, though, it is highly plausible that these truths are a priori entailed by nonmodal truths: in particular, by 'Water is H₂O' and 'Hesperus is Phosphorus'. Given that the relevant expressions are rigid (*de jure*), sentences such as 'If Hesperus is Phosphorus, it is necessary that Hesperus is Phosphorus' is a priori. So long as the relevant nonmodal truths are scrutable from *PQTI*, the modal truths are scrutable as well.¹⁰

¹⁰Chalmers (2012), p. 234.

On the appropriate scrutability thesis, the connection between the a priori and the necessary would be something like the following: Φ is a priori scrutable (from an appropriate base) iff Φ is necessary. The right-to-left direction is pretty clearly a consequence of a scrutability thesis. The left-to-right direction is slightly more complicated, especially if Φ contains any sort of vague predicates, partial descriptions of objects, or similarly incomplete notions. If this is the case, Φ might turn out to be some statement like “The sky is blue”. Given the notion of scrutability at work, it’s not hard to see that this turns out to be (in our sense) scrutable, but it is not in any obvious sense *necessary*,

One way one could solve this problem would be to ensure that Φ is described in such a way to ensure that its modal scope is explicit. For instance, the statement “The sky is blue” has a certain amount of relativity due to the utterer. When I utter “The sky is blue” in the actual world, I mean (presumably) that the sky at the actual world is (what I refer to and experience as) blue. In order to avoid this sort of complication, one could adopt a rigidification clause: each statement Φ is replaced in the base belief set of \mathcal{I} by some appropriate sentence (or proposition or non-linguistic object or...) Φ^p which specifies the worlds at which Φ holds. For instance, if Φ represents the statement “The sky is blue”, Φ^p would represent the statement “At worlds w_1, w_2, \dots, w_n , the sky is blue”.¹¹

¹¹With, of course, the requisite notion of “blue”, the proviso that the statement is indexed to a particular time (to allow for the possibility that the colour of the sky may vary over time), and other stipulations that would make the statement come out true in the actual world.

This seems to mitigate the difficulty encountered earlier. Should a statement of the type Φ^p turn out to be scrutable a priori, it seems quite plausible that it should turn out to be necessary. When we speak of a possible world, we are implicitly making the assumption that the world we speak of is maximally complete—that is, we refer to the totality of facts true in the particular world w whenever we say something like “ x is true at w ”. Now suppose that a statement like “possibly, x is not true at w ” were true, while simultaneously the statement “ x is true at w ” is true. Then there is some world w_1 at which the statement “ x is not true at w ” is true. But there is no accessibility relation R such that Rww_1 , since given the facts of w , it is not true that “ x is not true at w ”. Hence there is no such world w_1 accessible to w . The upshot of this is that if it is a priori scrutable that a Φ^p -type statement is true, then it is true in the actual world, and thus at all possible worlds. This gives us the other end of the biconditional.

A parallel version of this thesis cashed out in the language of ideal negative conceivability would be something like the following: Φ^p is conceivable_{in} iff Φ^p is necessary. Concomitantly, we can say that $\neg\Phi$ is not conceivable_{in} a priori iff $\neg\Phi^p$ is not necessary—or, in other words, if Φ^p is possible. So our equivalence would look something like $\neg C\Phi^p \leftrightarrow \Diamond\Phi^p$. This gets us the conditional $\Diamond\Phi^p \rightarrow \neg\Phi^p$, which in turn gets us the conditional $\neg\neg C\Phi^p \rightarrow \neg\Diamond\Phi^p$, which is our desired thesis.¹²

¹²Working classically and not intuitionistically, it should be said.

Should one want to find a counterexample to (–1.3), it is incumbent upon one to produce a case where something whose description is properly rigidified is not conceivable_{in} but is nonetheless possible.

What would it mean to produce such a counterexample? It would involve producing a statement which is determinably *false* a priori, but is in fact possibly true. At face value, one might be tempted to think that a statement similar to the ones in the previous section are what we are seeking. For instance, consider the statement “a particle can fail to have a definite location”; on the surface, that looks to be false a priori, and yet also true. This criticism loses its edge when we expand the statement out to the following: “given the formalism of quantum mechanics, it is coherent to suppose that the position of a particle can be described in terms of a superposition of possible positions.” Since our ideal conceiver has a maximally descriptive base belief set (a PQTI-complete one, to adopt Chalmers’ terminology for a moment), he will no doubt be able to conform the latter statement to his true beliefs concerning quantum mechanics, and so the problem would seem to be dissolved.

There seems to be a problem distinct from this one, and that is that different—entirely accurate—descriptions of the same entity (or state of affairs or...) may very well yield differing judgments concerning their conceivability_{in} or inconceivability_{in}. Suppose, for instance, that I ask our ideal conceiver to tell me whether or not he can conceive_{in} of the state of affairs accurately described by the statement “a system of sub-atomic particles, accurately described

by the correct laws of quantum mechanics, having a determinate position without being observed.” Supposing that the standard account of quantum mechanics is true, he will likely tell me that it is in fact inconceivable_{in} that such a situation should arise. From this, accepting (–1.3), we ought to conclude that such a state of affairs is not possible.

Now suppose I were to ask him to conceive_{in} of the state of affairs accurately described by the statement “a monocotyledon during the late Cretaceous having a determinate position without being measured.” What statement would he return? My suspicion is that he will return the answer that such a state of affairs is in fact conceivable_{in}. The problem here lies in the fact that the two statements in fact can accurately describe the *self-same* state of affairs. The monocotyledon, after all, is composed of proteins and other biological components, which in turn are composed of chemical elements, which in turn are certain types of atoms which contain certain types and numbers of sub-atomic particles, which are subject to the laws of quantum mechanics. The result, then, is that our ideal conceiver calls the same state of affairs conceivable_{in} and inconceivable_{in}—and thus, both possible and impossible.

This is similar to an arrangement offered by David Albert. The proposed experiment involves a “black” electron going through

a measuring device for “hardness,” and a competent¹³ observer named Martha:

[...]It follows from the linearity of the dynamical equations of motion (if those equations are right), and from what it means to be a competent observer of the position of the pointer, that the state when Martha’s done is with certainty going to be

$$(4.7) \quad \frac{1}{\sqrt{2}} |“hard”\rangle_o |“hard”\rangle_m |hard\rangle_e + \frac{1}{\sqrt{2}} |“soft”\rangle_o |“soft”\rangle_m |soft\rangle_e$$

That’s what the dynamics entails.

And of course what the postulate of *collapse* entails is that when Martha’s all done, then

$$(4.8) \quad \textit{either} |“hard”\rangle_o |“hard”\rangle_m |hard\rangle_e \textit{ (with probability } \frac{1}{2} \textit{)}$$

$$\textit{or} |“soft”\rangle_o |“soft”\rangle_m |soft\rangle_e \textit{ (with probability } \frac{1}{2} \textit{)}$$

is going to obtain.

And (4.7) and (4.8) are empirically different. The state described in (4.8) is the one that’s right; (4.7) is unspeakably strange. The state described in (4.7) is at odds with what we know of ourselves by *direct introspection*[...] *it’s a state in which there is no matter of fact about whether or not Martha thinks the pointer is pointing in any particular direction.*¹⁴

The point Albert seems to be making is that there is a severe difference between our ordinary macroscopic experience and what is

¹³See Albert (1992), p. 77: “What it means for Martha to be a competent observer of the position of a pointer is that whenever Martha looks at a pointer that’s pointing to “hard,” she eventually comes to *believe* that the pointer is pointing to “hard”; and that whenever Martha looks at a pointer that’s pointing to “soft,” she eventually comes to believe that the pointer is pointing to “soft”.”

¹⁴ibid, pp. 77–79. I have changed the numerical designations a little for ease of reference.

entailed by one of our best physical theories (or at least by the standard interpretation of it). This isn't quite the point made above, but it is closely connected, because we can run a very similar argument as above here.

Suppose I ask our ideal conceiver to undertake the relevant mental process needed to determine whether (4.7) is a conceivable_{in} state of affairs. He pauses for a moment, looks up, and says that it is. Then I ask him whether (4.8) is a conceivable_{in} state of affairs. There are two possibilities here: either he says it is, or it isn't. (Obviously.)

Suppose he says that it is conceivable_{in}. Then we clearly have a problem, because the states of affairs described by (4.7) and (4.8) are not compatible ones. It isn't that neither of these states of affairs are *possible* when taken together in different world, but that they can't be jointly true in a world sufficiently like ours as to be described by the assumptions being made at the outset. In other words, it's certainly possible for (4.8) to obtain in a world that *doesn't* function according to the relevant dynamics Albert takes as operative assumptions. Likewise, it's possible for (4.7) to obtain in a world where the kind of introspection Albert talks about (and presumably the ideal conceiver would need) isn't possible. But these two are incompatible if taken to be true at the *same* world.

And that is precisely what we want to be true if we want to hold onto this notion. Our ideal conceiver has essentially *human* cognitive powers, even if greatly enhanced and equipped with a maximally and relevantly complete base belief set. And thus the

result that *should* be returned, if the ideal conceiver is using correct phenomenological assumptions and postulates about what that kind of procedure is *usually* like for humans, is (4.8). On the other hand, if he bases his conceiving_{in} on a comparison of the correct quantum-dynamical approach (assuming, of course, that our quantum theory is complete enough to allow us to say that the equations of motion and evolution of state represent the real world), he will come to the conclusion that (4.7) is correct. And these are mutually incompatible states of affairs.

This creates an obvious problem. *If* the ideal conceiver, when asked to conceive_{in} of the state of affairs that is a result of the thought experiment suggested by Albert, uses all the contradiction-detection and reasoning faculties at his disposal, he could arrive at one of four possible conclusions:

- (1) (4.7) is the correct representation and (4.8) is the false one.
- (2) (4.8) is the correct representation and (4.7) is the false one.
- (3) Both (4.7) and (4.8) are correct representations.
- (4) Neither (4.7) and (4.8) are correct representations.

Suppose (1) is correct. Then our regular introspective faculties are faulty for these sorts of situations. But it is hard to see just how one could maintain this and still retain the definition of an ideal conceiver that we've been working with. Certainly there are some states of affairs which our ideal conceiver will be able to conceive_{in} of by using just these phenomenological principles or correct descriptions (for instance, whether it is possible for an object to be both green and blue all over, under the correct phenomenal description

of both green and blue, and so forth). And we're going to want to be able to say that using these faculties to determine their possibility (or lack thereof) is a reliable heuristic. But it's hard to see how we can maintain this when they fail in domains where, by their own lights, they *should* be accurate. If using all faculties together leads to contradictory conclusions, that counts against the reliability of them in general.

Considering whether (2) is correct doesn't seem to bear good results either. It is just to jettison one set of primary conceivability heuristics for another. If our quantum-mechanical theories are mistaken in this arena and the correct phenomenal principles are the ones to be used, then we run into virtually the same problem as before, only with the positions reversed. And if we suppose that here are phenomenal laws which are not reducible to microphysical laws, and are in some deep sense *fundamental*, then it is hard to see which of these decisions we should respect. If two (supposedly fundamental) sets of laws yield inconsistent descriptions of a situation in which *both* would seem applicable, then we have a problem—unless there is some deeper set of laws to which we need to make recourse in order to settle the dispute. But postulating such a fudge is to trade one mystery for another; this is essentially the postulation of an oracle, which we forbade in our formulation of the ideal conceiver. So this isn't a very viable solution.

If (3) is correct, then we have a rank contradiction. These two states are mutually incompatible, and so (3) has to be inconceivable_{in}

as well if we've endowed our ideal conceiver with the law of non-contradiction.

If (4) is the correct representation, then both our quantum-dynamical theories *and* psychophysical or phenomenological principles—both of which we have what looks like very good reason to think are true—are fundamentally mistaken or flawed, in ways we have no knowledge of. It is possible that there is some unifying description that will render the end of the thought experiment intelligible to someone in possession of the correct theory, but that's not of much assistance to us. The person defending (\neg 1.3) can simply say “no, there is a correct description but we just don't know it yet,” but this is not a very convincing argument. Arguably, our confidence in our best quantum-dynamical theories and at least some tier of our phenomenological experiences should exceed our confidence in (\neg 1.3), since it's founded on them in the first place.¹⁵ Thus, it seems on the face of it more warranted to reject (\neg 1.3) than to reject either of them.

Maybe there's another way out. For instance, many physicalists would hold that our psychophysical and phenomenal lives are in the end reducible to the neurochemistry of particular brain states which accompany them. For them, each correct phenomenal principle would be an elliptical for an exhaustive (though in all likelihood, extremely un-economical) chemical or physical description of that particular brain state. For instance, one could hold that the

¹⁵Recall that the base belief set, if it's to be complete, is going to have to contain all the true beliefs that allow our ideal conceiver to reach these conclusions. So in some sense these principles are logically prior to (\neg 1.3).

phenomenological description of the particular principle in questions can be Ramseyfied to contain only terms which are either reducible to purely physical terms or else be themselves purely physical terms.¹⁶ Since the physical is fundamental, then, any conflict between what our phenomenal experience seems to tell us and what our best physical theories tell us will be resolved in favor of the physical theory.

The trouble with this is that, even if descriptions of phenomenal laws all have Ramsey sentences containing only physical terms, the phenomenal descriptions will still have to be *accurate*, even if only as a placeholder. Any inconsistency with our best physical theories that occurs as a correct application of the phenomenal laws will *itself* have to come as a result of a conflict between differing parts of our best physical theories, which is hardly a less daunting problem. Physicalists or functionalists tend to be queasy about relying on introspection as any sort of infallible guide, so one might be tempted to say that the apparent antinomy is a result of one of those failures. But simply denying that introspection is *always* reliable isn't enough to win the argument, since its general reliability is not in question. One has to instead say that, in situations of this particular type, introspection is unreliable. Simply using this as a one-off example of its unreliability is a bit too close to special pleading for comfort. But this is a situation in which most of us think that introspection is valid, since—on a purely pragmatic level—it yields good results in other cases. For instance: I make a free throw during a game of

¹⁶*A la* David Lewis, in [Lewis \(1972\)](#).

basketball, and my introspective judgment while making the shot tells me that the result will be that either the ball will go through the hoop or it will not; and, every time I have made that judgment, the disjunction has turned out to represent the outcome accurately. So any argument to the contrary has to shoulder the burden of explaining why *this* sort of introspection is in general reliable, but in this case not reliable.

A point of clarification is in order. In the preceding discussion, the reader may have gotten the impression that I'm taking the position that observation by some consciousness-bearing entity is essential to wave-function collapse. That is, in order for there to be a fact-of-the-matter (in the way we ordinarily like to think of it) about how the world is, there has to be some conscious observer to observe (however we want to understand that term) the evolution of a wave-function and thus to collapse it. This might evoke a certain amount of incredulity in the reader, and still more in the proponent of the inconceivability_{in} rule. "Surely," he says, a mischievous glint in his eye, "if you're allowed to use some sort of conscious observer, existent whenever there is wave-function collapse, I'm allowed to use an ideal conceiver with abilities beyond the mere human in order to circumvent the problems you raise!"

This would be both an understandable and—to my mind—not entirely incorrect response, if indeed this argument were being made. Luckily, it is not. It does not require or assume that there is some sort of necessarily phenomenal or conscious component to wave-function collapse. All that is assumed is that there is, in the way

we ordinarily understand it, a fact-of-the-matter about, say, the answer that Martha's going to give. And *this* in turn—or so the claim goes—requires wave-function collapse. Now certainly, the way in which we think that wave-function collapse normally occurs is in fact when a conscious observer makes a measurement of a given quantum-mechanical system.¹⁷ No particular assumptions are made, nor any doctrines advanced, concerning what precisely causes wave-function collapse, and certainly not the quasi-idealistic doctrine that *conscious observers* are needed for that to take place. All that is required is the following: if there is to be a fact-of-the-matter about physical systems or macroscopic objects more generally, wave-function collapse (provided that this is the correct interpretation of quantum mechanics) should have taken place.¹⁸

Of course, it may turn out that the orthodox interpretation of quantum mechanics is false. Perhaps the pilot-wave or many-worlds interpretations are correct. The argument offered above is not (quite) predicated on the truth of the orthodox version. Rather,

¹⁷Though strictly speaking, I think one might make do by simply saying that what is required is that a measurement of the system is somehow made, since—as far as I understand it—the conscious registering of the particular result of the measurement is not what determines that result, but rather the measurement itself.

¹⁸I say “fact-of-the-matter,” and this may induce the reader to think I mean *determinate* fact-of-the-matter, in the sense of saying of a given system, “this and only this will happen.” What I have in mind is a little softer, and not so proscriptive. It is more like the following: for a system (under a suitably broad interpretation of that word, encompassing physics, day-to-day experience, and so on) to admit of a fact-of-the-matter, it must have some *definite*, though not necessarily *determinate*, configuration. This admits of, for instance, disjunctive facts-of-the-matter: either Chelsea will win the match, or Manchester will win it. The fact-of-the-matter of the outcome of the soccer match (prior to its conclusion) is not determinate in the sense of not allowing for a certain outcome, but is *definite*, in the following sense: one of the two teams *will* win the match, assuming draws are disallowed. There will be no “state mixture.”

it maintains the following: we have physical models of the world which conflict in certain ways with some theoretical accounts of modal knowledge. If we are being good scientific realists, then we will need to take this tension seriously. This does not necessarily amount to a rejection of the modal theory, but it does highlight the need for a reassessment.

This argument used the idea of wave-function collapse and quantum-mechanical systems in an attempt at precisification and illustration. It is possible, however, that the reader has taken that to be the sum total of the argument. This would be an incomplete notion, since the class of problems envisioned goes a certain degree further. The idea is, instead, that it seems like certain states of affairs, where this phrase is taken in its most broad interpretation, may be accurately and irreducibly portrayed in terms of conflicting and (apparently) mutually exclusive descriptions.

To get at what this means, it may be helpful to examine a slightly more prosaic example. Let us borrow an example from Miss Anscombe:

Kant introduces the idea of “legislating for oneself,” which is as absurd as if in these days, when majority votes command great respect, one were to call each reflective decision a man made a vote resulting in a majority, which as a matter of proportion is overwhelming, for it is always 1-0. The concept of legislation requires superior power in the legislator. His own rigoristic convictions on the subject of lying were so intense that it never occurred to him that a lie could be relevantly described as anything but just a lie (e.g. as “a lie in such-and-such circumstances”). His rule about universalizable maxims is

useless without stipulations as to what shall count as a relevant description of an action with a view to constructing a maxim about it.¹⁹

The idea is that, under the Kantian system, one is left without a guide to what precisely will cast the vote as to the relevancy—or indeed the *accuracy*—of a certain description of the action in question. Certainly one might very well give assent to the statement “lying is wrong,” but one would likely find it a little more difficult to agree with the statement “lying to the SS concerning the Jews in your basement, with an eye towards protecting the innocent, is wrong.” But one may, without sacrificing a whit of *accuracy* or even *relevancy*, say that the description of “lying” or “lying towards a good end” fits the situation entirely accurately. And the point is that the system has a hard time telling you, without being able to make reference to consequences, just which of these descriptions is the one you should use when trying to see whether your maxim might be a universal law. So we get an inconsistency, because the state of affairs is—irreducibly and accurately—described in terms where either maxim applies. And introducing a stop-gap measure to solve this *particular* case is not going to be good, because these sorts of rules are supposed to be generally applicable.

Cases of intentional action are also intriguing here. Miss Anscombe has another illustrative example of this peculiarity: suppose a bird lands on a twig to pick at some food on the twig. This twig also happens to be smeared with bird-lime (a substance used for trapping birds). But “[t]he bird wanted to land on the twig alright,

¹⁹Anscombe (1958), p. 2.

but did not want to land on a twig smeared with bird-lime.”²⁰ The moral we should draw from the example is that one can, without loss of accuracy, describe the same state of affairs—the bird landing on just this or that twig—by two descriptions. First, “the bird intended to land on that twig with the bird-seed,” and second, “the bird intended to land on that twig with the bird-lime.” If I ask \mathcal{I} whether he can conceive_{in} of the state of affairs under both descriptions, my suspicion is that he would answer “yes” to the former. Accompanying that suspicion is one which says that he would find the second somewhat more difficult. To the degree that birds are capable of intentional action, and to the degree that they are capable of rational examination, it would be quite odd to treat the bird (rationally) taking an ultimately detrimental action. The case is one of incomplete knowledge, to be sure, but the paradox remains: the state of affairs becomes perfectly conceivable_{in} under one (completely accurate) description, and somewhat more opaque under another (equally accurate) description.

A possible—and interesting—response to this view might be to treat statements of the form “ x intends to ϕ ” as not admitting of free substitutions of ϕ . To treat statements such as “The bird intends to ϕ ” as admitting of liberally substituting different (true) descriptions of an action ϕ would, on this view, be akin to entering into a “second grade of intentional involvement.” Substituting in different descriptions of ϕ will change the intension of the statement, and consequently its truth value. Say Λ is a statement operator standing

²⁰Anscombe (1982), p. 209.

for “intends to”. Then statements of the form $\Lambda x\phi$ for an agent x and an action ϕ will only admit of substitutions for ϕ whenever the new description of x ’s action is a description reflecting some part of his ϕ -ing that he knows he does.²¹

So cases involving statements of the type $\Lambda x\phi$ seem to be more problematic. This doesn’t seem to be so in the case of the righteous liar, however, since it involved no such statement operation. All that we are required to analyze are the descriptions of the lie, one with reference to the action, one with reference to the intention of the action. But here the question is not whether the liar intended to lie, else the problem would not arise. Hence, the antecedent problem—that of asserting the liar intended an action he did not—is not of particularly serious consequence here.

The general case of this puzzle goes roughly as follows. Some states of affairs are describable in multiple ways. Some states of affairs are even describable in multiple, equally—and irreducibly—accurate ways. And some of these simultaneously accurate descriptions are mutually incompatible—in Anscombe’s case, “lying” and “lying to protect the innocent”; in Albert’s case, the everyday phenomenal description and the quantum-mechanical description. Single-instance fudges, which tell you which description is correct in a *single* case are not useful, because they set up larger conflicts between two supposedly fundamental and correct sets of generally-applicable principles. And general rules of thumb—e.g., “not lying always wins,” “physics trumps phenomenology”—are

²¹This response runs along the lines of Anscombe’s own analysis of such situations; see pp. 12-13 of [Anscombe \(1963\)](#).

problematical, because they don't explain the (at the very least apparent) applicability of the set of maxims and/or principles which are deemed to *not* apply, which seems to cut against their inclusion in the "fundamental" category. This, in the end, leads to what we might characterize as "antinomies of pure conception."

3. The upshot

If the previous sections hold, then the notion that inconceivability leads to impossibility is somewhat fraught. How is this relevant to RR? It doesn't immediately imply that it's false, since RR is logically equivalent, not to the inconceivability thesis, but to the thesis that *impossibility* implies inconceivability.

There's another worry. In the case of (\neg 1.3), we say that P is inconceivable if an ideal conceiver can rule P out on purely a priori grounds. Now, if P is ruled out on these grounds, this means that the falsehood of P is determined on those grounds, which entails that the truth of $\neg P$ is a priori. The claim is, further, that if $\neg P$ is a priori, P cannot be. This should not be difficult to believe. After all, if one denies that P is not a priori, then one is left in the situation of believing both that $\neg P$ is and that P might be a priori.

If P is inconceivable, P is not a priori. This doesn't seem momentous. But realize for a moment what this would mean. The state of affairs "P is not a priori" is implied by the state of affairs "P is inconceivable," but it is also *on the face of it* compatible with the situation of P's being conceivable. For all the phrase "P is conceivable" means in this context is that P is not ruled out a priori. But if

we suppose that P is both not ruled out a priori and that P is not a priori—which is really just another way of saying that neither P nor $\neg P$ are a priori—there isn't any apparent contradiction, which there very much *should* be if P is both inconceivable and conceivable. One way of getting around the dilemma would be to maintain that P 's not being a priori in fact implies that P is ruled out a priori. But there seem to be some problems with just such a position. For one thing, this would preclude there being any true proposition whose truth is not determinable a priori, since—say—if a proposition was determinably true a posteriori but not a priori, it would not be true a priori, and consequently it would be ruled out a priori. This would leave one in the situation of having a proposition whose negation is true a priori but which is also true a posteriori. This is—it should go without saying—a tough pill to swallow, since there seem to be any number of propositions—those concerning correct legal judgments, say—whose truth is determinable in this way.

The claim isn't that this constitutes a thorough refutation of the two theses presented so far, because in order to do that one would need to offer a much more thorough argument than the one presented. In fact, there's a very simple way of avoiding the apparent paradox: if one were to maintain, that for every proposition P , either P or $\neg P$ is a priori, the problem disappears quite easily. This is distinct from saying that P is not determinable a posteriori, as we did above, since it doesn't seem impossible for the truth of a given proposition might be determinable both a priori and a posteriori. For instance, one might learn the truth of the proposition "The

basketball will fall when thrown” either by extrapolation from Einstein’s field equations, knowledge of what the noun “basketball” means, and exhaustive knowledge of the initial conditions of the ball/player system, or by observing, as did our forefathers throughout most of history, that stuff falls downwards when not held up (barring very unusual situations, e.g., parabolic flight paths).

Rather, the point is this: if we think that conceivability of a certain kind *does* indicate—in fact, entail—possibility, then one would expect—I claim—expect for there not to be situations in which there’s an apparent compatibility between P’s being conceivable and P’s being inconceivable (in the appropriate way). This would be expected because the relationship is supposed to be one of entailment, rather than reliability. If the relation were like that between, say, our perception of the eternal world and the actual external world (one of reliability), any apparent compatibility between situations which have some contradiction or inconsistency in their perceptually-relevant elements wouldn’t refute the reliability. All it would indicate is that the apparatus can produce mistaken results—and as we don’t think that such paradoxes mean our senses are de facto unreliable²², we wouldn’t think that such inconsistencies would mean conceivability would be totally unreliable.

On the other hand, if we think that there’s some sort of more structural relation between conceivability and possibility of the kind we’ve been talking about, we’d expect that there are none of these sorts of compatibilities; the structure would preclude them. But,

²²At least, the non-sceptics amongst us do.

as argued above, there do in fact seem to be such compatibilities, which is the opposite of the expected result. This doesn't mean the position is roundly refuted. Instead, it gives *good reason* to believe either that the position *as such* is false or else the position *as formulated* is false.

The argument has, at present, been against the inference from conceivability to possibility, but the consequences of the mentioned inquietudes extend a bit further. Take, for example, the thesis (sometimes called "modal rationalism") that possible worlds just are mental representations; more specifically, that the necessary is the a priori. Keith Hossack, for example, has it that "According to modal rationalism, the necessary is the a priori. If that is correct, we can give the following definitions: a fact is necessary if it is a priori, contingent."²³ Suppose P is a proposition whose negation is a priori. It follows then that P is necessarily false, and hence is true in no possible world. This very neatly comes within the purview of the arguments we've been making—certain kinds of P-worlds are in fact possible (actual, even), but if we put on our modal rationalist hats for a moment, we'd be led to conclude that they aren't.

What's a nice modal rationalist to do? Supposing the offered arguments are successful, there are at least two options. One is to give up modal rationalism wholesale; the other is clarify precisely *which* a priori propositions are necessary. Hossack constructs the relevant notion of apriority as follows:

²³Hossack (2007), p. 438.

We begin by defining the primitive: a fact is primitive if in favourable circumstances a person can know the fact just by taking thought.[...] We define the a priori in terms of the primitive, as follows. First, we stipulate that every primitive fact is a priori. Next, we stipulate that all and only the logical consequences of a priori facts are a priori also. Finally we define logical consequence in terms of the primitive. Say a relation R is a consequence relation if it is a structural relation with the following property: it is primitive that whenever some propositions Σ stand in R to a proposition P , then if Σ are all true, then P is true also.²⁴

I suppose that one could restrict the “favorable circumstances” in some way that would disallow the multiple-descriptions problems arising from using different sorts of descriptions (or other types of such arguments): via some tie-breaking clause between the conflicting principles or laws of description, to give one example. But it’s not clear how this would avoid the problems that we encountered in the previous example. On the other hand, one could say that any state of affairs which yields the kind of propositions that accurately describe it in a mutually-inconsistent way are not, in fact, a priori; thus one avoids the problem by denying it. But all this does is eliminate P as an a priori proposition; P is contingent if P is not a priori, but there’s still the problem that P yields mutually inconsistent yet accurate descriptions. If I were to represent a P -world as “a world w at which all accurate descriptions of P are true,” all I would be doing is making the inconsistencies implicit. All worlds containing inconsistent yet actualized propositions contain contradictions, and

²⁴ibid., p. 439.

these are impossible worlds. So we still get $\neg P$, necessarily, and we're back where we started.

One might be concerned, on the other hand, that the denial of the coextensivity between the conceivable and the possible commits one to a sort of modal dualism. On this view, metaphysical modality and epistemic modality are two fundamentally different sorts of things. And this—or so the worry goes—is an un-parsimonious view.

Take Frank Jackson, for instance:

I have two reasons for holding that there is only one sense of necessity and possibility in play here. The first is Occamist. We should not multiply senses of necessity beyond necessity. The phenomena of the necessary a posteriori, and of essential properties, can be explained in terms of one unitary notion of a set of possible worlds. The phenomena do not call for a multiplication of senses of possibility and necessity, and in particular for a distinction among the possible worlds between the metaphysically possible ones and the conceptually possible ones.²⁵

So what's a good Ockhamist to do? Certainly it is true that it is more ontologically parsimonious to deal only with one class of possible worlds (which requires identifying the epistemically possible and the metaphysically possible). But if the argument given above goes through, then the identification of the two commits one to an even stranger ontological category: that of worlds with contradictory features. If there are inconceivable worlds which are also possible (actual, even), then either the identification is ill-advised or else needs to be amended to allow for worlds which are both conceptually possible (i.e., conceivable) and conceptually

²⁵Jackson (2000), p. 70.

impossible (i.e., inconceivable).²⁶ And that is a queer metaphysic indeed.

A further question needs to be addressed: how should this affect confidence in RR? Suppose that, as argued above, there are some propositions whose conceivability seems to be compatible with their inconceivability. It's intuitively appealing to think that this produces a tension in the rationalist view. But just *how much* of a tension does it produce, and just how should we then react to the success of the argument? It's hard to know how to proceed here without endorsing a particular notion of just how belief-defeating works, and what should be the proper response to a general sort of defeasibility argument. And that is not the purpose of this present essay. Instead, the argument offered is methodological, one founded on a general principle of theoretical compatibility.

We often think of predictive success as a hallmark of an effective explanation of a phenomenon (or of a theory). One way of explicating this notion is something like the following: if P is a group of phenomena in need of an explanation, and τ is a theory which is supposed to explain P, then τ should predict—that is, have amongst its consequences—all elements of P.²⁷ There is a related notion of predictive success: keeping τ and P as before, τ should not be consistent with the falsity of any of the constituent phenomena of P.

²⁶For a fuller discussion of these sorts of arguments—though not one strictly along the lines we're pursuing—see Chapter 1 of [Kment \(2014\)](#).

²⁷I concede that “supposed to explain” is a very vague phrase. I here mean it in the highest level of generality, meant to encompass causal, grounding, and any other kind of explanation. The notion of predictability is, to me, most comfortably used in the realm of the physical sciences, but I don't wish to parochialize it to my personal interests.

Note here that, in such fields as physics, typically more phenomena will be added to P in the course of experiments conducted to test the predictions of τ .

As a thorough guide to the use of prediction, this is obviously underdeveloped. But we're not at the moment concerned with general scientific methodology. We are instead concerned with examining a method for ascertaining how we should evaluate the degree to which the predictions of a theory ought to affect our commitment to the theory. Further, this is put forth, not so much as a principle to which one ought adhere no matter what, but rather a heuristic. All else being equal, it seems better to prefer theories whose predictions or consequences are consistent with the phenomena they wish to explain, and inconsistent with the opposite phenomena.²⁸ And, all else being equal, if we're confronted with a theory whose implications are consistent with the falsity of some of its other predictions, this seems to count against this theory's optimality.

Why is this? The gist is that if a theory implies that conceptually contradictory things are consistent, or have certain features which are consistent, this suggests either a fault in the methods by which the consequences are ascertained or else a fault in the theory itself. This, in turn, suggests that we ought to amend either (or perhaps both) of these. In this case, we ought to reject or amend either the

²⁸"Opposite phenomena" might seem a little opaque when speaking of the natural sciences, since not all predictions have a neat and tidy "opposite." With regards to philosophical topics, however, one can get more straightforward opposites: a state of affairs being conceivable and a state of affairs being inconceivable, for instance.

methods used to reach the paradoxical compatibility (that is, my argument) or the theory we started out with (the royal road).

So why should we accept this sort of heuristic? We should do so because we use this sort of reasoning quite often already, and to good effect. More to the point, we use it to good effect in everyday life. Suppose I judge that it's raining because I heard a resounding crash while I was inside studying. On the basis of this, I judge that the ground outside should be wet. Now, suppose I actually go outside in about a half-hour and observe that the sun is high in the sky, the ground is dry, and it isn't actually raining outside at that time. Now, certainly, that result is, loosely speaking, *consistent* with my hypothesis that it was raining. It could be that I in fact did hear thunder, but that that only occurred towards the end of the thunder-shower, and the sun came out immediately after and dried the ground.

Suppose I say to myself "well, that doesn't mean that I'm wrong." Technically I'm right, because those things *are* both consistent with my theory that it's raining. The point is that this compatibility should either give me reason to doubt the line of reasoning that led me to conclude that the actual situation is consistent with my theory, or else to conclude that the explanation I've given of the thunderous noise (that is was raining) is in fact faulty.

The suggestion is that we use this sort of heuristic all the time, or at least think we should. Whenever a rational person is faced with data which conflicts with some pre-existing theoretical commitments, he will often re-examine those commitments, to see if

they conflict after all. And if his theory is consistent with two mutually exclusive scenarios, that should give him just the slightest bit of pause, and perhaps lead him to change the views he holds. Such is—the argument runs—the case with modal monism, and hence with RR. The fact that the theory is consistent with the conceivability *and* the inconceivability of the scenarios we've been discussing in the preceding paragraphs should, all else being equal, give us doubt as to whether RR is a useful—never mind valid—method for obtaining knowledge of possibility or necessity.

Closing Remarks

As the reader may have gathered from the previous chapter, I do not think the unrestricted use of conceivability to access possibility is methodologically sound. In this sense, then, there are inconceivable worlds: worlds which contain aspects which are either irreducibly inconceivable or conceivable under one (accurate) description and inconceivable under another (equally accurate) description. But I do not think that these worlds are strange or terribly peculiar—indeed, they may go so far as to be wholly familiar.

What use, then, for conceivability? Should it be discarded as useless, or left unanalyzed as having no philosophical import. The latter is certainly not my view. Just the opposite: I have spent over fifty pages under the assumption of its falsehood!

Is it useless? I do not think so, though the reader may have gotten that impression from the preceding pages. Rather, my view is this: the *unrestricted* use of conceivability to discover what is possible is, in fact, illegitimate. But there may yet be many legitimate uses for a philosophically elaborated theory of conceivability.

On the level of phenomenology, for instance, conceivability is quite interesting. What *precisely* do we do when we conceive_O of something? What are the characteristic mental processes or activities involved? These are non-trivial and potentially complex

questions. And such questions are, or at any rate ought to be, of interest to any philosopher.

Further, while conceivability might be a poor way to discover possibility, it might be a good way to determine *compossibility*. While conceivability_O is subject to our abilities and concepts, and so is a poor guide to possibility *tout court*, it may be a good guide to determining what things are consistent with our concepts. Similarly (and more generally), conceivability_n might be a good way to determine the joint consequences of philosophical positions, much in the way thought experiments are supposed to do. It might even be possible to construct a somewhat formal and consistent procedure using conceivability_n which would allow thought experiments to be conducted and assessed in fairly uniform ways.

Whatever the philosophical merits of conceivability, I do not think that “being a good guide to determining possibility” is among them. It might be a good guide to what might be true *given* the beliefs and knowledge of a particular person. But as a guide to anything more than highly parochial, agent-bound seeming possibilities, it seems to be hopelessly flawed. Taken as having some close connection to possibility, we arrive (I have argued) at the conclusion that we live in either an impossible world or a world whose possibility is indeterminate. And this is a paradox. So far from being able to conclude that inconceivable worlds are impossible, we are led to conclude, instead, that our world—horses, sea horses, glaciers, supermassive black holes, dark matter, gamma ray bursts, and things

our philosophy, natural or speculative, may not yet dream of—is
inconceivable.

Bibliography

- Albert, David (1992), *Quantum Mechanics and Experience*, Harvard University Press, Cambridge, Mass.
- Anscombe, G. E. M. (1958), "Modern Moral Philosophy", *Philosophy*, 33, 124, pages 1-19.
- (1963), *Intention*, Harvard University Press, Cambridge, Mass.
- (1982), "Under a Description", in *The Collected Papers of G. E. M. Anscombe, vol. II: Metaphysics and the Philosophy of Mind*, Wiley Blackwell, Oxford, chapter 19, pages 208-219.
- Bealer, George (2002), "Modal Epistemology and the Rationalist Renaissance", in *Conceivability and Possibility*, edited by Tamar Szabo Gendler and John Hawthorne, Oxford University Press, Oxford, chapter 1, pages 72-125.
- Chalmers, David (2002), "Does Conceivability Entail Possibility?", in *Conceivability and Possibility*, edited by Tamar Szabo Gendler and John Hawthorne, Oxford University Press, Oxford, chapter 3, pages 145-200.
- (2004), "Epistemic Two-Dimensional Semantics", *Philosophical Studies*, 118, pages 153-226.
- (2012), *Constructing the World*, Oxford University Press, Oxford.
- Geirsson, Heimir (2005), "Conceivability and Defeasible Modal Justification", *Philosophical Studies*, 122 (3 2005), pages 279-304.

- Hossack, Keith (2007), "Actuality and Modal Rationalism", *Proceedings of the Aristotelian Society*, 107, pages 433-456.
- Jackson, Frank (2000), *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford University Press, Oxford.
- Jarvis, Jonathan Ichikawa and Benjamin (2012), "Rational Imagination and Modal Knowledge", *Nous*, 46, 1, pages 127-158.
- Jenkins, C. S. (2010), "Concepts, Experience, and Modal Knowledge", *Philosophical Perspectives*, 24, pages 255-279.
- Kant, Immanuel (2003), "The Only Possible Argument", in *Theoretical Philosophy, 1755-1770*, edited by David Walford, Cambridge University Press, Cambridge, chapter 5, pages 107-201.
- Kment, Boris (2014), *Modality and Explanatory Reasoning*, Oxford University Press, Oxford.
- Kripke, Saul (1980), *Naming and Necessity*, Harvard University Press, Cambridge, Mass.
- Lewis, David (1972), "Psychophysical and Theoretical Identification", *Australasian Journal of Philosophy*, 50, 3, pages 249-258.
- Menzies, Peter (1998), "Possibility and Conceivability: A Response-Dependent Account of Their Connections", *European Review of Philosophy*, 3, pages 255-277.
- Murphy, Peter (2006), "Reliability Connections Between Conceivability and Possibility", *Dialectica*, 60 (2 2006), pages 195-205.
- Peacocke, Christopher (1997), "Metaphysical Necessity: Understanding, Truth, and Epistemology", *Mind*, 106, pages 521-574.
- (1998), *Being Known*, Oxford University Press, Oxford.

- (2002), “The Principle-Based Account of Modality: Elucidations and Resources”, *Philosophy and Phenomenological Research*, 64, 3, pages 663-679.
- Priest, Graham (1998), “What is So Bad about Contradictions?”, *The Journal of Philosophy*, 95, 8, pages 410-426.
- Putnam, Hilary (1975), “The Meaning of ‘Meaning’”, *Minnesota Studies in the Philosophy of Science*, 7, pages 131-193.
- Reid, Thomas (1941), *Essays on the Intellectual Powers of Man*, Macmillan, St. Martin’s Street, London.
- Resnick, Robert (1968), *Introduction to Special Relativity*, John Wiley and Sons, New York.
- Roca-Royes, Sonia (2011a), “Conceivability and de re Modal Knowledge”, *Nous*, 45 (1 2011), pages 24-49.
- (2011b), “Modal Knowledge and Counterfactual Knowledge”, *Logique et Analyse*, 54, pages 537-552.
- Vaidya, Anand (2008), “Modal Rationalism and Modal Monism”, *Erkenntnis*, 68, 2, pages 191-212.
- Van Inwagen, Peter (1998), “Modal Epistemology”, *Philosophical Studies*, 92, 1, pages 67-84.
- Williamson, Timothy (2007), “Philosophical Knowledge and Knowledge of Counterfactuals”, *Grazer Philosophische Studien*, 74, 1, pages 89-123.
- Worley, Sara (2003), “Conceivability, Possibility, and Physicalism”, *Analysis*, 63 (1 2003), pages 15-23.
- Yablo, Stephen (1993), “Is Conceivability a Guide to Possibility?”, *Philosophy and Phenomenological Research*, 53 (3 1993), pages 1-41.