RELATIONAL POSSIBILITY

Daniel Berntson

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY

THE DEPARTMENT OF PHILOSOPHY

Advisors: Shamik Dasgupta and Boris Kment

September 2019

Abstract

This dissertation consists of three papers. The first is about relational possibilities, which are possibilities that compare or otherwise relate things across worlds. We might say, for example, that Socrates could have been taller than he is or that the Athenians could have been happier than they are. The standard view is that such claims require quantification over further things like heights or degrees of happiness. But as we will see, this approach stands in the way of an especially promising strategy for doing science with only particles. In response, I develop and defend an alternative view that I call modal relationalism. According to the modal relationalist, modality is ultimately about how things could have *differed*, not just how things could have been, and so can naturally accommodate the needed comparisons without quantifying over further things.

The second paper presents a paradox: Counterfactuals are somewhat tolerant. Had Socrates been at least six feet tall, he need not have been *exactly* six feet tall. He might have been a little taller. He might have been six-foot one or six-foot two. But while counterfactuals are somewhat tolerant, they are also bounded. Had Socrates been at least six feet tall, he would not have been a thousand feet tall, for example. Surprisingly, given these simple assumptions, we can prove a flat contradiction using principles validated by our best semantic theories. These include the familiar similarity analysis from David Lewis.

After sketching the paradox, I describe what I think is the solution.

The last paper is a kind of technical companion to the first. It describes a hierarchy of multi-dimensional quantified modal languages that are modeled using a corresponding hierarchy of multi-dimensional Kripke models. We then show how to build multi-dimensional proof systems and prove completeness. This is relevant to the first paper because, if modal relationalism is true, then we are going to need an appropriate multi-dimensional modal language. This third paper shows that such languages are in good working order from a certain technical perspective.

Acknowledgements

They say that it takes a village to raise a child, and the same goes for the writing of a dissertation. In my own case, I am especially grateful for the many sharp and stimulating conversations with my advisors Shamik Dasgupta and Boris Kment. Without their help and constant encouragement, I would not be the philosopher—nor the human being—that I am today. I have also had immensely helpful conversations with many other friends and colleagues, including Nat Tabris, Gideon Rosen, Mark Johnston, Hans Halvorson, and John Burgess.

My wife Skye-lindey has lived with these ideas from the start. She has often had the misfortune of being the first to hear new arguments, but has always responded to them with the same charity and grace. The ideas in this dissertation exist in their present form only because we worked on them together—sometimes in the car, sometimes at the dinner table, and sometimes late at night after putting our daughter Alice to bed.

That I am a philosopher at all is thanks to my parents Joanne and Steven. Growing up, we had long conversations about big ideas—conversations that I would later learn were called "philosophy". I never had to convince them that going to graduate school was worth doing. It was just a way of continuing what we had always done together as a family. I could never thank them enough for their many years of love and encouragement.

Finally, I would like to thank those who attended talks in which early versions of these papers were presented. These include talks given at the University of Toronto, the Fourth Taiwan Philosophical Logic Colloquium, the University of Victoria, the Slovak Academy of Sciences, and the Princeton University Philosophical Society.

Contents

1 Relational Possibility

Absolute possibilities are about how things are at individual worlds. We might say that Socrates could have been a farmer or that Athens could have defeated Sparta. Relational possibilities are about how things compare or otherwise relate across worlds. We might say that Socrates could have been taller than he is or, perhaps, that the Athenians could have been happier than they are.

Relational possibilities, while ordinary and familiar, turn out to be surprisingly hard to express systematically. The usual solution is to express them by quantifying over things like heights or degrees of happiness. As it turns out though, this natural solution has certain unexpected implications. One of them is that science is committed to far more ontology—to a wider variety of things—than we might otherwise have thought. The fundamental physical world will have to include not just particles, but things like numbers, spacetime points, or distances.

I think we can do better. My aim in this paper is to convince you that we can understand relational possibilities without quantifying over further things. Not only does this better reflect our ordinary thinking about modality, it also gives us a powerful philosophical tool, one that can help us do science with only particles.

1.1. Expression

We have two ways of talking about possibilities. One is by quantifying over things like merely possible worlds or merely possible individuals. We might write $\exists w Fsw$ to say that there is a possible world at which Socrates is a farmer or $\exists x(Cxs \land Fx)$ to say that there is a possible counterpart of Socrates who is a farmer. The other is by using modal operators instead. In that case, we would write $\Diamond Fs$ to say that it could have been that Socrates was a farmer.

Corresponding to these two ways of talking, there are competing ideas about the nature of possibility. One is that possibility is ultimately about certain *things*—either possible worlds or possible individuals. The other is that possibility is ultimately about a certain *mode* or *manner* in which conditions are satisfied. The first view says that Socrates could have been a farmer *because* there is a possible world at which he is a farmer. The second says that no, on the contrary, the direction of explanation is reversed. There may be a possible world at which Socrates is a farmer, but that is only *because* Socrates could have been a farmer.

Questions about relational possibility look somewhat different depending on whether we take worlds or modal operators to be more basic. For now, we are going to consider the matter from the perspective on which operators are most basic.

Suppose we have a language with a taller-than predicate, names for individuals, and a possibility operator. This lets us say that Aristotle is taller than Socrates by writing Tas. We can then say that Aristotle could have been taller than Socrates by writing $\Diamond Tas$. Now we want to express the relational possibility of Socrates having been taller than he actually is. How should we do that? We could try $\Diamond Tss$. But this says that Socrates could have been taller than himself, not that Socrates could have been taller than he is. We could try adding an actuality operator to our language. This might seem promising, since it lets us talk about the actual world within the scope of a possibility operator. We can then write $\Diamond @Tss$. But this is true only if Socrates is *actually* taller than himself— which of course he is not. So it would seem that we have simply run out of syntactic combination. There is no way to say that Socrates could have been taller than he actually is. Call this the **problem of expression**.

One idea for solving the problem can be traced to Bernierand Russell's paper "On Denoting" (1905). Russell is there interested in understanding certain belief attributions. You might, on seeing a friend's new yacht, report that you had thought it would be larger than it is. Your friend is a bit touchy and so says no, obviously not. The yacht is *exactly* as long as it is. The joke gets its punchline because the target sentence is ambiguous between two readings. Russell thinks that what you intended to say is that the size you thought the yacht was is greater than the size the yacht is. Or in other words:

> There is a unique size y such that the yacht has y and you believed that (there is a unique size x the yacht has and $x > y$).

The reading your friend attributes to you is the one on which the belief operator takes wide scope. That reading is of course absurd—no sane person would believe that the size of the yacht is greater than the size of the yacht.

The same strategy can be used to express relational possibilities. Suppose we replace belief with possibility and sizes with heights and the yacht with Socrates. We then have:

> There is a height y such that Socrates has y and $\Diamond$(there were a height x such that Socrates had x and x were greater than y).

This would seem to be a perfectly good way of expressing the idea that Socrates could have been taller than he is. If we let the possibility operator take wide scope, we get the false claim that Socrates could have been taller than himself. This solution obviously generalizes beyond sizes and heights. As we proceed, then, let's refer to things like sizes and heights as **degrees** and to Russell's strategy as the **degree solution**.[1]

So far so good. We can express certain simple relational possibilities using the degree solution. While this is certainly progress, the solution faces three basic problems. The first is that the original claims would not seem to involve such quantification. We never *said* that there is a height that Socrates has. What we *said* was that Socrates could have been taller than he is. For all we said, there may be no heights. The existence of heights may be impossible.

In a way, using height comparisons as our stock examples can obscure this basic point. After all, heights are pretty ordinary sorts of things. We have singular terms for them and we do, on occasion, quantify over them in ordinary conversation. Consider other sorts of comparisons though. Suppose that after seeing your friend's yacht, you say to yourself:

> My friend could have been less offended than he was.

In order to say that using the degree solution, we have to quantify over degrees of

---

1. Variations of the degree solution have also been endorsed by Morton (1984), Cresswell (1990), Milne (1992), and Kemp (2000).

offendedness. But did you really *say* that there are degrees of offendedness? Surely not. Maybe there are no degrees of offendedness because offendedness comparisons are too vague or otherwise ill-behaved to support the existence of such things. Still, it would seem, what you said about your friend is true. He could have been less offended than he was. What you said is simply *neutral* with respect to the question of whether or not there are degrees of offendedness.

The second problem is that while quantification over degrees lets us express certain relational possibilities, there are still others that cannot be expressed even with such quantification. For example:

> It could have been that every tree on the quad were taller than it is.

There is no way to say that even *with* quantification over heights. The best we can do is:

> For every tree on the quad x, there is a height y such that x has y and $\Diamond$(there were a height z such that x had z and z were greater than y).

But this is not strong enough. What we want to say is that the trees on the quad could have *collectively* been taller than they are. But this only says that the trees on the quad could have *individually* been taller than they are. We could try adding an actuality operator, but this is no help either.[2] The sentences is still inexpressible. The degree solution, then, is at

2. Using an actuality operator, the best we can do is:

> $\Diamond$(For any x such that @(x is a tree on the quad), there were heights y and z such that @(x has y) and x had z and z were greater than y).

This is pretty close. In terms of worlds, this says that there is a possible world w such that for every x in w... What we want, though, is a sentences that says that there is a possible world w such that every x in the actual best only a partial solution to the problem of expression.

Maybe these are oddities we could learn to live with. Maybe we could learn to live with the idea that, despite all appearances, when we say that Socrates could have been taller than he is, we are saying that heights exist. Maybe on reflection, certain more complex relational possibilities make no sense, and so there is no need to express them. Still, we face a third problem, which is that the degree solution comes with certain strong metaphysical requirements. One of those requirements is that things like heights exist—we need *there to be* a height that Socrates has such that he could have had a greater one. The mere existence of heights, though, is not enough. Heights must also have a certain character. Even if you accept that there are heights, you might not be able to use the degree solution.

When it comes to a quantity like height, we can distinguish two views. Absolutism says that height is ultimately about the determinate heights things have. Socrates has a

---

world v... The problem, in other words, is that the variable x ranges over the domain of the possible world instead of the domain of the actual world.

It turns out that the sentence *is* expressible if we have quantification over heights *and* an actuality operator *and* plural quantification:

> The xx include all and only the trees on the quad and $\Diamond$(the xx existed and that for every x among the xx, there were a height y such that x had y and @(there is a height z such that x has z and y is greater than z)).

But now we have a sentences that is a mismatch for the original in two ways. We not only have quantification over heights, where the original sentence has none, but plural quantification over trees, where the original sentences has only individual quantification over trees. That these sorts of syntactic gymnastics are required to express a relatively simple possibility suggests that we have not found the real logical form of the sentence.

height of six feet and this, the absolutist says, is among the most basic facts about height. Comparativists say no, on the contrary, height is ultimately about certain comparative relations like *being taller* and *being the same height*. The basic facts about height describe a web of such relations between individuals.[3]

The degree solution requires not just the existence of height, but the truth of absolutism. Here is why: Some comparativists deny flat-out that there are heights—our talk about Socrates having a height of six feet is just a confused remnant of discredited common sense. More likely, though, a comparativist will want to explain heights using height comparisons. For example, she might say that the height of Aristotle is identical to the height of Socrates because Aristotle is the same height as Socrates. Or she might say that the height of Socrates is six feet because he is the same height as certain standard reference objects—standard six foot measuring rods or something similar.

Now, if she is going to use the degree solution, the comparativist needs to explain de re modal facts about heights. Why could Aristotle have had the very height that Socrates actually has? The most straightforward answer is that the identity of heights across worlds works the same as the identity of heights across individuals. Aristotle could have had a height identical to the height Socrates actually has because Aristotle could have been as tall as Socrates actually is. But this means that we are explaining de re modal facts about heights *using relational possibilities*. But in that case, we cannot very well go back and do the reverse. We cannot *also* explain relational possibilities using de re modal facts about heights,

3. See (Dasgupta 2014) for further discussion.

as the degree solution would have us do. So the degree solution requires absolutism.[4]

Putting the matter more generally, the degree solution requires degrees to not just exist, but to be *prior* to relational possibilities. One implication is that the solution is not available to comparativists. Another, which we will consider in the next section, is that the solution prevents us from doing science with minimal ontology.

1.2. Science with Minimal Ontology

Our best scientific theories describe a world built out of certain things, like quarks and bosons, that have certain features. We thus naturally think of science as having ontological commitments. Insofar as we believe our best theories, we are committed to believing in the things they describe—we are committed to believing in quarks and bosons and such. The question then is, how far do those commitments extend? What are the ontological commitments of science?

Science as we know it tells us about various physical quantities like mass, charge, and distance. Those quantities are described using numbers. To fix on an example, suppose we perform a series of experiments and discover that the movement of particles is described by

4. You could imagine a view on which the height of Aristotle at w is identical to the height of Socrates at v because Aristotle occupies the same position in the network of height relations at w as Socrates does at v. We would then be free to use heights to explain relational possibilities, since they are not presupposed. In that case, though, the appeal to heights is redundant. We can just say that Aristotle at w is the same height as Socrates at v because he occupies the same position in the network of height relations. So more carefully, we might say that the degree solution is *useful* only if absolutism is true.

Newton's laws of motion. These laws require distance ratios between particles to determine how they move.[5] What are distance ratios? Suppose we use a meter stick to determine that *ab* are two meters apart and *bc* are one meter apart. We could then record the result using a distance ratio function from particles to numbers by writing $\delta(abbc) = 2$. If you like, think of this as a certain definite description:

> The distance ratio of *ab* to $bc = 2$.

Such measurements let us apply the laws and predict how things move.

How should we understand such distance ratios? One view is that we should take science at face value. Distance ratios ultimately involve a relation between particles and numbers and, so, numbers play an essential role in the physical world. Science is as much committed to numbers as it is to quarks and bosons. A competing view is that numbers are useful but not essential. They are part of science only because they speed along certain reasoning. With enough time and patience, we could fully describe the physical world without them. There may be physical facts involving numbers but, if so, they will have to

---

5. Here is a simple case. Suppose the world is Newtonian with gravity the only force. There are three particles *abc* with *b* between *ac*. The particles are at rest relative to one another and *a* is as massive as *bc* put together. Then, the collision of *a* and *b* will be simultaneous with the collision of *b* and *c* just in case the distance ratio of *ab* to *bc* is $\sqrt{2}$. If there is no determinate distance ratio, the laws will fail to determine whether those collisions will be simultaneously.

be ultimately explained using physical facts not involving numbers.[6,7,8]

Suppose we agree that numbers are not essential to the physical world. We then face the challenge of showing how there can be distance ratios without numbers. One strategy for doing that requires nothing more than particles and a pair of spatial relations. Those relations are betweenness and congruence. Intuitively speaking, betweenness is the relation of one thing being on a straight line between two others. Distance congruence is the relation of two things being exactly as far apart as two others. Now imagine a world with exactly

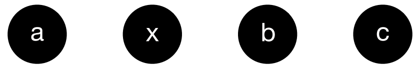
---

6. The first view is endorsed by Quine (1961). Field (1980) defends the second.

7. As Field (1984) points out, there are different sorts of reasons you might think that numbers are dispensable. You might be a nominalist who denies flat-out that numbers exist. Alternatively, you might accept the existence of numbers, but deny that there are any basic physical relations between physical things and numbers. Field calls this latter view moderate platonism. Nominalists, then, need to show why scientific claims involving numbers are useful without being true. Moderate platonists need to show why they are true using only physical facts that do not involve numbers.

8. For reasons of space, we are not going to spend much time either (a) presenting arguments for nominalism or (b) survey existing nominalist views in the literature. For those interested, these issues are pursued at length in (Burgess and Rosen 1997). Besides considerations of space, though, there are also philosophical reasons for not pursuing these matters. Namely: The importance of doing science without numbers simply does not hang on the question of nominalism. Even if I were fully prepared to accept the existence of numbers, I would still want an *intrinsic* characterization of the physical world that did without them. Moreover, many of the nominalist strategies in the literature would not give us an appropriately intrinsic characterization. For example, one common nominalists strategy is to use quantification over *numerals* in place of quantification over numbers. But numerals, it seems to me, are if anything *less* likely to play an essential role in the physical world than numbers. So even if numbers could be replaced with numerals, it would be no help.

four point particles that looks like this:



The betweenness and congruence relations of the world are as they appear. The particle x is between ab, the particles ax are congruent with xb, and so on. It also looks like the ratio of the distance between particles ab and bc is two, and in fact they are. What we want to do is explain this distance ratio in terms of the basic congruence and betweenness relations.

Just looking at the illustration, you can basically see how the explanation goes. Each adjacent pair of particles is congruent with every other adjacent pair, so we can treat those pairs as "units". There are then two "units" between ab, but only one "unit" between bc. So ab are twice as far apart as bc. A bit more carefully, say that x is halfway between ab when x is between ab and ax and xb are congruent. We then claim that ab are twice as far apart as bc because:

> There is an x halfway between ab such that ax and bc are congruent.

Other rational distance ratios can be defined similarly. Irrational distance ratios are defined using limits.

All of this would seem to work beautifully. The problem is that the proposed reduction works only if there happen to be enough particles and they happen to be in the right place. Suppose we have a world just like the last except that we delete the second particle.

This is a world in which *ab* is twice as far apart as *bc*. Our proposed definition of that distance ratio, though, requires there to be something halfway between *ab*. Since there is nothing there, the definition fails.[9]

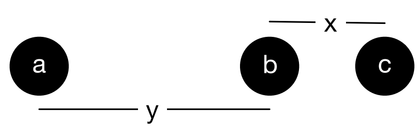
The problem can be fixed if we accept the existence of spacetime points. Spacetime points, like particles, stand in betweenness and congruence relations. Unlike particles, though, spacetime points are highly organized—you can always count on them to be where they need to be. In particular, there will be laws guaranteeing that whenever there are two things, there is a spacetime point halfway between them. Adding spacetime points to our three particles world, then, we have a world with a spacetime point *x* halfway between *ab*.

Because there is once again *something* halfway between *ab*, our proposed definition of the distance ratio works just fine.[10]

9. You might wonder whether this is just a failure of imagination. Sure, *one* proposed reduction to betweenness and congruence fails, but maybe some other reduction could succeed? In fact, it could not. For any real number *r*, the distance ratio of *ab* to *bc* could have been *r*. So there are uncountably many ways to configure *abc* in terms of distance ratios. But there are only *finitely* many ways to configure *abc* in terms of betweenness and congruence relations. So there is no way to reduce the distance ratio configurations of *abc* to the betweenness and congruence configurations of *abc*.

10. Early attempts to axiomatize Euclidean space using betweenness and congruence include (Veblen 1904) and (Pieri 1908). The project was later advanced by Alfred Tarski and his students, who give increasingly simple axiom schemes in (Tarski 1952), (Tarski 1959), and (Gupta 1965). The observation that distance ratios would seem to require quantification over either numbers or spacetime points was originally made by

Another strategy would be to accept the existence of distances rather than spacetime points. What are distances? There are different views you could have, but perhaps the simplest is that distances are first-order universals.[11] We then have two important relations. The first is a three-place addition relation between distances. This is just the relation of distances *xy* together being as long as *z*. The second is a having relation between pairs of particles and distances. To see how this all works, consider a world in which there are three particles, with *ab* three times as far apart as *bc*.



Here, particles *ab* have distance *y* and *bc* have distance *x*. We can then easily explain why *ab* is twice as far apart as *bc*. That is because there are distance *x* and *y* such that *x* and *x* add up to *y*, with *ab* having *y* and *bc* having *x*.[12]

We can explain distance ratios without numbers, then, if we quantify over either

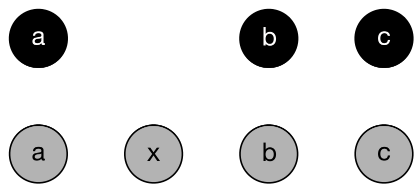
Field (1984).

11. Brent Mundy (1987) describes a higher-order variation of this view. For him, distances are binary relations instead of first-order predicables. He then has second-order quantification over these relations and a second-order addition relation between them. Myself, I prefer the first-order version. I can see no advantage that would justify the added cost of second-order ideology.

12. It should also be pointed out that the proposed strategy requires not just the existence of distances, but the existence of *uninstantiated* distances. Just consider a case like the one above, except that *ab* are three times as far apart as *bc* instead of twice as far apart. In that case, there will have to be an *x* between *bc* such that there is a *y* and there is *z* between *ab* such that *x* and *x* sum to *y* and *x* and *y* sum to *z*. But so long as *abc* are the only particles, *y* will be uninstantiated. This means that Aristoteleans about distances cannot make

spacetime points or distances. You might wonder, though, whether these are the only options. Can we explain distance ratios using only particles? Or does science inevitably require further things?

The natural strategy at this point would be to give a modal theory of distance ratios. To see how that goes, consider the following diagram. The top row of black particles represents how things actually are. The bottom row of grey particles represents how things could have been.



What we want to explain is why the particles *ab* are twice as far apart as *bc* on the top row, since the top row represents how things actually are. The modal solution is to say that *ab* are twice as far apart as *bc*, not because there is actually a particle halfway between *bc*, but because there *could* have been. Particles could have been the way they are on the bottom row. There is thus no need for further ontology. Rather than accepting that there *are* numbers, spacetime points, or platonic universals, we only need to accept that there *could* have been more particles.

The problem is that we have now run headlong into the problem of expression. In order for the modal solution to work, we need to describe our diagram using modal operators.

---

use of the proposed strategy, for example, since they are willing to quantify over instantiated universals, but not uninstantiated universals.

We need to say that:

> It could have been that *a* were as far from *b* as *a* is from *b* and that *b* were as far from *c* as *b* is from *c* and that there were an *x* such that *x* were between *a* and *b* and *a* were as far from *x* as *x* were from *b* and *x* were as far from *b* as *b* were from *c*.

But the requisite possibility is a *relational* possibility. We need to be able to say that it could have been that the actual particles *were* as far apart as they *are*... The only way to say that in language of modal logic is to use the degree solution. We have to quantifying over further things like numbers, spacetimes points, or distances. But quantifying over such things is exactly what we are trying to avoid! So it looks like we have reached the end of the line. Science would seem to require more than just particles after all.

Now in fact, I think we have not reached the end of the line. What we need to do science is not more ontology. What we need is a better solution to the problem of expression.

1.3. Grammatical Mood

Natural language has two very different systems for making modal claims. On the one hand, we have quantification over worlds. We say that *there is* a world in which Socrates is a farmer or that *every world* is one in which Socrates is human. On the other hand, we have modal auxiliaries. We say that Socrate *could* have been a farmer or that he *must* have been human.

The modal auxiliary system is generally regimented with **standard modalese**. This is just the language of standard quantified modal logic. It has quantifiers, variables, names, predicates, and truth functional connectives. Besides these, it has a sentential possibility operator. This operator is meant to play the role of the modal auxiliary 'could'. $\Diamond Fs$ says that it *could* have been that Socrates were a farmer. We then define the necessity operator $\Box = \neg\Diamond\neg$, which is meant to play the role of 'must'. $\Box Hs$ says that it *must* be that Socrates is human.

The problem is that while standard modalese can express some of what we say with ordinary modal auxiliaries, it cannot express everything. We can say more with modal auxiliaries than we can with modal operators.

In natural language, we make modal claims using a combination of modal auxiliaries and **grammatical mood**. When talking about how things are, we use the indicative mood. We say that Socrates *is* a philosopher or that the Athenians *are* wealthy. When talking about how things could have been, we use the subjunctive mood. We say that it could have been that Socrates *were* a farmer of that it could have been that the Athenians *were* happy.[13]

13. Over time, English has undergone a long process of simplification. The result is that distinct subjunctive morphology has all but disappeared, when it used to be much more common. Verbs in the subjunctive mood generally have the same morphology as verbs in the simple past tense. The lone exception is the verb 'to be'. Strictly speaking, the first and third person subjunctive form of 'to be' is 'were'. The first and third person simple past form is 'was'. But because the process of simplification continues apace, it can often sound just as natural—or even more natural—to use 'was' in place of 'were'. We might say that it could have been that Socrates *were* a farmer. Or we might say that it could have been that Socrates *was* a farmer. In both cases, the verb 'to be' is in the subjunctive mood. It's just that in the second case, we are using simple past morphology to express the subjunctive mood. As we go along, I will generally use 'were' instead of 'was' as the first and third person subjunctive form of 'to be'. But if you find that this sounds stilted or even ungrammatical, try using

All of this is important because in natural language, we use certain patterns of grammatical mood to express relational possibilities. Suppose we start with the claim:

It could have been that Aristotle were taller than Socrates were.

This sentence expresses an absolute possibility. It is true iff there is a possible world such that Aristotle at *that* world is taller than Socrates at *that* world. In terms of syntax, the key features is that this sentence has a predicate with two copulas, both of which are in the subjunctive mood. Now watch what happens when we put the second copula in the indicative mood:

It could have been that Aristotle were taller than Socrates is.

Now we have a sentence that expresses a relational possibility. It is true iff there is a possible world such that Aristotle at *that* world is taller than Socrates at *our* world. We could emphasize the intended reading by saying that it could have been that Aristotle were taller than Socrates *actually* is. But semantically speaking, this is not required.[14]

Standard modalese can express the first sentence without quantifying over heights, but not the second. This is bizarre, at least on the face of it. Both sentences would seem to be built out of a modal operator, a two-place predicate, and a pair of names. The only

'was' in place of 'were'. Note also that what I am calling the subjunctive mood is really the *past* subjunctive mood. English also has what is called the *present* subjunctive mood. We use the present subjunctive when we say, for example, that we should finish this footnote, lest it *be* longer than it already is.

14. See (Lewis 2001, pp.13-14), (Wehmeier 2003), and (Mackay 2013) for more on the role of grammatical mood in natural language.

difference is the mood the second copula. But why should *that* make any difference? Why should saying 'is' instead of 'were' require further ontology?

Maybe we could learn to live with such oddities. The more serious problem, though, is that we can use grammatical mood to build more complex sentences that are inexpressible even with quantification over heights. We saw this phenomena earlier:

> It could have been that every tree on the quad <u>were</u> taller than it <u>is</u>.

From the perspective of standard modalese, this perfectly ordinary claim is nothing but gibberish—mere sound and fury signifying nothing. But the ordinary claim is not gibberish. The ordinary claim is perfectly sensible.

As far as doing science with only particles, you may have noticed that the possibilities we need are themselves perfectly expressible in natural language. All we need are modal auxiliaries and the right pattern of grammatical mood. One example is the sentence at the end of section two. Another is the first axiom of the modal theory of distance ratios that we are going to give later on. That axiom has the form:

> It could have been that for any two particles, those particles <u>were</u> as far apart as they <u>are</u> and...

Standard modalese can express relational possibilities only if we quantify over things like distances. But ordinarily, we can express them without such quantification. Just look at the above sentence! Nowhere do we say that there are distances. Moreover, even if we were to quantify over distances, this particular sentence would *still* be inexpressible.[15] The

15. The intended reading is the one on which it is true iff there is a possible world w such that for any two

problem, then, is standard modalese, not the sort of possibilities we need to do science with only particles. Those possibilities are perfectly sensible.

1.4. Relational Modalese

I think the problem has both a clear diagnosis and a clear solution. Standard modalese has modal operators, which play the role of modal auxiliaries, but nothing that plays the role of grammatical mood. It is not surprising, then, that we can say more with natural language than we can with standard modalese. What we need, then, is a modal language that *does* have something that plays the role of grammatical mood. Once we do, we should have a much better match for natural language.

The question then is how to do that. How do we build a modal language that has something playing the role of grammatical mood? There are different ways to go here. The language I favor is **relational modalese**. Like standard modalese, relational modalese extends the language of quantified predicate logic by adding modal operators. It has a possibility operator $\Diamond$ that can be used to define a necessity operator $\Box = \neg\Diamond\neg$. These play the role of the modal auxiliaries 'could' and 'must'. The difference is that relational modalese also has **mood operators** that play the role of grammatical mood. These operators $\uparrow$ and $\otimes$ are called up and swap respectively. We can then use them to define a third $\downarrow = \otimes\uparrow$ called down.

particles in the actual world v, those particles at w are as far apart as they are at v. Since they have the same general form, this sentence is inexpressible for the same reason that the sentence in the last paragraph is inexpressible. Like that sentences, it is also inexpressible even if we add an actuality operator.

We have a new language then. But how does it work? Here is the basic idea. Suppose we have a taller than predicate Txy. When not in the scope of a modal operator, we read the predicate as being uniformly in the indicative mood.

Txy $\quad$ x is taller than y is.

When inside the scope of a modal operator, on the other hand, we read the predicate as having mixed mood.

$\Diamond Txy$ $\quad$ It could have been that x were taller than y is.

We can then use mood operators to permute the grammatical mood of the copulas.

$\Diamond \otimes Txy$ $\quad$ It could have been that x is taller than y were.

$\Diamond \uparrow Txy$ $\quad$ It could have been that x were taller than y were.

$\Diamond \downarrow Txy$ $\quad$ It could have been that x is taller than y is.

As you can see, swap exchanges the moods. Down copies the mood of the second copula to the first. Up copies the mood of the first copula to the second.[16]

16. Most predicates have only a single copula—take the 'is a farmer' predicate for example. In terms of relational modalese, think of those predicates as having both a pronounced mood and an unpronounced mood that is "stored" for future use. The stored mood then plays the role that the second copula plays when a predicate has a second copula. So for example, $\Diamond Fx$ says that it could have been that Socrates were a farmer. The predicate is in "mixed mood" in the sense that the subjunctive mood is pronounced and the indicative mood is stored. Quantifiers can be thought of in the same way. $\Diamond \exists x$... says that it could have been that there *were* an x... The quantifier is in mixed mood in the sense that the subjunctive mood is pronounced and the indicative mood is stored. So to give a couple of example, $\Diamond \exists x \otimes Fx$ says that it could have been that there *were* an x such that x *is* a farmer. $\Diamond \otimes \exists x \otimes Fx$ says that it could have been there *is* an x such that x *were* a farmer.

While the above translations are completely ordinary, they can also be a little hard to parse. It is usually helpful, then, to also see how relational modalese can be translated into talk about worlds. In terms of worlds, sentences without modal operators describe the actual world.

Txy	x at our world is taller than y at our world.

When in the scope of a modal operator, predicates then describe relations across worlds by default.

$\Diamond Txy$	There is a possible world such that x at that world is taller than y at our world.

We then use mood operators to permute which of the two worlds we are talking about within the scope of a modal operator.

$\Diamond \otimes Txy$	There is a possible world such that x at our world is taller than y at that world.
$\Diamond \uparrow Txy$	There is a possible world such that x at that world is taller than y at that world.
$\Diamond \downarrow Txy$	There is a possible world such that x at our world is taller than y at our world.

Swap lets us exchanges the two worlds. Up lets us make a claim purely about the first world. Down lets us a make a claim purely about the second.

We should point out that these translations in terms of worlds have a somewhat

---

Anyway, you get the idea.

different status than the translations in terms of modal auxiliaries. The translations in terms of modal auxiliaries are correct in the sense that they have *the same meaning* as the corresponding sentences in relational modalese. Relational modalese is a language for expressing *the very same thoughts* that we ordinarily express with modal auxiliaries. The translations in terms of worlds, on the other hand, are only correct in the sense that they have *the same truth conditions*. If they also had the same meaning, the claim that Socrates could have been a farmer would have the same meaning as the claim that there is a possible world at which Socrates is a farmer. There would thus be no sensible question about whether Socrates could have been a farmer *because* there is a possible world at which he is a farmer, or vice versa. But that is a sensible question. So the translations in terms of worlds do not have the same meaning. That said, they are still quite useful for getting fix on how the new operators work.

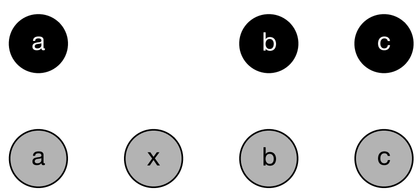
Now you might wonder, why are we using sentential operators to play the role of grammatical mood? After all, natural language has no such operators. There are no 'it is swap that' or 'it is up that' operators. The short answer is that this is the simplest way to replicate what we do in natural language without reworking the basic syntax of predicate logic. In natural language, mood is a matter of conjugating *verbs*. But predicate logic has no verbs. It only has predicates. There is thus no possibility of directly replicating what we do in natural language without reworking the basic syntax of predicate logic. Maybe that can—and even should—be done.[17] A simpler solution, though, is to use mood operators.

17. I have some sympathy for this project myself. From a historical perspective, the Fregean idea that the *predicate* should be a basic unit of syntax is something of an anomaly. The medievals generally followed

These may not be a perfect match for the syntax of natural language. But as you can see, such operators can still play the same semantic role as grammatical mood. And that is what really matters, or so it seems to me.

We are going to make all of this more precise later on. You get the basic idea, though, for how relational modalese works. Relational modalese is just a formal device for doing *what you already know how to do* in natural language. You already know how to use modal auxiliaries and mood to express relational possibilities without quantifying over further things. Because relational modalese has not just modal operators, but mood operators, you can do the same with relational modalese.

This lets us solve the problem of expression without quantifying over things like heights. It also puts us in a position to do science with only particles. For present purposes, doing science with only particles means describing the diagram from the end of section two.



The top row represents how things are. The bottom row represents how things could have been. We now claim that *ab* are twice as far apart as *bc* because a certain modal fact obtains.

Aristotle in thinking that sentences were built by linking terms for predicables and terms for individuals with a copula. Maybe what the problem of expression shows, then, is that Aristotle was right and Frege wrong. Modern modal logic has a problem of expression because the whole modern logical enterprise started off on the wrong syntactic foot, so to speak. See (Parsons 2014) for more on the role of the copula in medieval logic.

$$\Diamond(Cabab \wedge Cbcbc \wedge \exists x{\uparrow}(Baxb \wedge Caxxb \wedge Cxbbc))$$

This is sentence of relational modalese. But how does it work? The sentence is built using a pair of basic predicates, which are betweenness and congruence.

Babc	*b* is between *a* and *c*
Cabcd	*a* is as far from *b* as *c* is from *d*

Predicates have mixed mood by default when in the scope of a modal operator. So

$$\Diamond(Cabab \wedge Cbcbc \wedge \ldots)$$

says that it could have been that *a* were as far from *b* as *a* is from *b* and *b* were as far from *c* as *b* is from *c*. The existential quantifier is in the subjunctive mood by default when in the scope of a modal operator. So

$$\Diamond(\ldots\exists x\ldots)$$

says that it could have been that...there were an *x* such that... The up operator puts everything that follows into the subjunctive mood. So

$$\Diamond(\ldots{\uparrow}(Baxb \wedge Caxxb \wedge Cxbbc))$$

says that it could have been that...*x* were between *a* and *b* and *a* and *b* were as far apart as *x* and *b* were and *x* and *b* were as far apart as *b* and *c* were. Putting it all together:

> It could have been that *a* were as far from *b* as *a* is from *b* and that *b* were as far from *c* as *b* is from *c* and that there were an *x* such that *x* were between *a* and *b* and *a* were as far from *x* as *x* were from *b* and *x* were as far from *b* as *b* were from *c*.

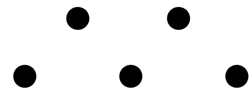
This is exactly the sentence from the end of section two that we wanted to express.[18] So we have arrived. We can explain distance ratios using only particles.

All sorts of details need to be filled in—and will be filled in section six and the appendix. Still, you can see the big picture. Relational modalese better reflects ordinary modal thought and talk. As a result, it can express relational possibilities without quantifying over further things. This gives us a powerful tool for reducing our ontological commitments. We can, for example, explain distance ratios without appealing to numbers, spacetime points, or distances. The same basic strategy naturally extends to other physical quantities like mass. We are thus well on our way to doing science with minimal ontology.

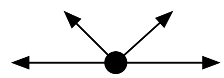
1.5. Modal Relationalism

We have all been told, at one time or another, that possibility is about how things could have been. This is the view more or less built into standard modalese. Possibilities are about how things could have been in themselves and can be fully described independently of how things are. When I think of possibility as ways things could have been, I imagine our world as one of many dots in modal space. Like this:

18. In terms of worlds, this sentence can be translated as follows: Let the actual world be v. There is then a possible world w such that ab at w are as far apart as ab are at v and bc at w are as far apart as bc are at v and there is an x in w such that x is between ab at w and such that ax at w are as far apart as xb at w and xb at w are as far apart at bc at w.



Relational modalese ultimately represents not just a different way of talking about possibility, but a different way of thinking about its nature. Possibilities are ultimately not ways things could have been, but ways things could have *differed*. When I think about possibility this way, I imagine vectors with our world as the origin.



Rather than being points that can be independently characterized, possibilities are something like departures or displacements from reality. Call the first view **modal absolutism** and the second **modal relationalism**.

What we have seen is that given the first picture, relational possibilities come with ontological commitments. On the second, we can do without those commitments. The question then is, why did we find the first picture so compelling in the first place?

Let's try a pair of experiments. First, close your eyes. Imagine that Holmes and Dr. Watson are sitting in their favorite easy chairs at 221 Baker Street. Holmes is holding up his thumb and long, slightly crooked index finger. His fingers are about three inches apart and exactly halfway between them is a blue marble suspended in midair. "Isn't that curious," he tells Dr. Watson.

Now the second. Hold your thumb and index finger two or three inches apart and focus on the point halfway between them. Imagine there is a blue marble suspended there, spinning slowly and reflecting the light. Stare at that point for awhile and imagine as much detail as you can. Can you see it? I find that I can almost see a ghostly hologram.

These two experiments represent two ways of imagining—two modes of conceivability if you will. In the first case, you close your eyes and build a world *ex nihilo*. In the second, you leave your eyes open and "edit" the actual world in various ways. You hold certain features fixed, like the distance between your fingers, while applying certain transformations. You might add a blue marble here, delete the red pen over there, or double size of your coffee mug. Where the first kind of conceivability is a kind of virtual reality, the second is a kind of augmented reality, you might say.

This matters because we philosophers, I think, tend to imagine worlds in the first way. This tends to encourage the idea that possibility is about ways things could have been. After all, when you close your eyes and build a world form scratch, the actual world goes away. There would thus seem to be no way to imagine basic comparisons *between* the actual world and the world you construct in thought. By contrast, on the second way of imaging worlds, the actual world does not go away. You can just *see* that the world you are imagining is one in which your fingers are as far apart as they actually are. And so it would seem to be a mode of conception that allows for basic comparisons across worlds.

Another reason we might tend to think of possibilities as ways things could have been is that tend to think of them as being "self-contained" in the same way the actual world is self-contained. More carefully, you might think that the actual world can be fully characterized without appealing to merely possible worlds. But the actual world is one among equals. The only difference between our world and merely possible worlds is that our world happens to be actual. That means that just as the actual world can be fully characterized without comparing it to merely possible worlds, merely possible worlds can be fully characterized without comparing them to either the actual world or each other. So there are no basic comparisons across worlds.

This little argument has two major premises, both of which can be resisted. The first says that the actual world can be fully characterized without appeal to other worlds. But if modal notions play a fundamental role in the physical worlds—as I maintain they do—then this simply is not so. We may need comparisons across worlds to determine actual distance ratios, for example. The second premise says that if the actual world can be characterized without appeal to other worlds, then so can other possible worlds. But again, this need not be. Those of us inclined towards actualism generally think of the actual world as playing a special role in modal space. Part of that special role may be that the actual world can be fully characterized without comparing it to merely possible worlds, even though merely possible worlds cannot be fully characterized without comparing them to the actual world.

A third reason that we may tend to think about modality in absolute terms is that we tend to think about possible worlds as something like maximal consistent sets of sentences. A sentences ϕ is then true at a world w iff ϕ is a member of w. But in that case, possible worlds are fully characterized by their intrinsic features, since truth at a world is an intrinsic matter.

While there is something to be said for the idea that possible worlds are sets of sentences, it should be emphasized that this is a bit of theory, not a part of ordinary modal thought. For example, ordinarily speaking, we can just just say that:

There is a possible world at which Socrates is taller than he is at the actual world.

Using talk about maximal consistent sets of sentences, on the other hand, we need to not just quantify over sets of sentences, we need to quantify over *heights*. We need to say something like:

> There are heights x and y and there are maximal consistent sets of sentences w and v such that the sentence saying that Socrates has x is a member of w and the sentence saying that Socrates has y is a member of v and the sentence saying that x is greater than y is a member of v, where the sentences that are members of v are precisely those sentences which are true.

But the ordinary sentence has no such quantification. Thinking of worlds as maximal consistent sets of sentences, then, results in a certain mismatch with our ordinary talk about worlds.[19]

---

19. Besides such difficulties, there are other reasons to think that possible worlds are not maximal consistent sets of sentences. For example, as Fine (2005) points out, you might think that just as any possible human could have been a human, any possible world could have been *the world*. But the world—the sum of everything—is not a maximal consistent set of sentences, nor could the world have been a maximal consistent set of sentences. So possible worlds are not maximal consistent sets of sentences.

There is also the notorious problem of accounting for worlds that differ only with respect to which merely possible individuals there are. For example, there is a possible world w in which there are exactly two merely possible bosons. One spontaneously decays and the other survives. But in that case, you might think, there should *also* be a possible world v in which the outcome is reversed. The merely possible boson that survives at w decays at v, and the merely possible boson that decays at w survives at v. But since merely possible bosons do not exists, they have no names, since they fail to stand in any denotation relations. And so w and v cannot be maximal consistent sets of sentences, since they are distinct, but there is no difference with respect to which sentences are true at those worlds. See (Nolan 2002) for more on this.

If possible worlds are not maximal consistent sets of sentences, then what are they? What I think, broadly speaking, is that talk about possible worlds is a roundabout way of talking about how things could have been. The right question, then, is not what possible worlds are. The right question is what makes talk about possible world *apt*. And the answer, I think, is that talk about possible worlds is apt when it appropriately corresponds to the modal facts, given in terms of operators. How this might work, exactly, will be illustrated in the next section.

1.6. Filling in the Details

In section four, we gave a partial characterization of relational modalese. We then used the language to explain a particular distance ratio without quantifying over things like numbers, distances, or spacetime points. In this section, we are going to do the same, but in a way that is both more general and more precise. The result will be a full axiomatic theory of distance ratios. That theory will explain distance ratios in *general* without quantifying over anything but particles. This will serve as a kind of proof of concept for using relational possibilities to do science with only particles.

The first tool we need to build our theory is a language for describing possible worlds. That language $\mathcal{P}$ will be called **possiblese**. $\mathcal{P}$ is a standard quantified predicate language with identity and five non-logical predicates:

Ww w is a possible world

Aw w is actual

Iaw a is in w

$Babcw$ b is between ac at w

$Cabwcdv$ a and b at w are as far apart as c and d at v

Using possiblese is straightforward. For example, suppose we want to say that there is a possible world in which there is a particle b between a and c. We can do that by writing:

$$\exists w(Ww \wedge \exists x(Ixw \wedge Babcw))$$

Notice that because the congruence predicate is indexed to a pair of worlds, we can describe basic congruence relations not only within worlds, but across worlds. We can say, for example, that there is a possible world at which a and b are as far apart as they are at the actual world:

$$\exists w(Ww \wedge \exists v(Av \wedge Cabwabv))$$

As you can see, sentences of possiblese are basically just direct transcriptions of what we say in natural language.

Our ultimate goal will be to give a theory of distance ratio using possiblese, then translate that theory into relational modalese. Where the original theory will have quantification over both worlds and particles, the translated theory will replace the quantification over worlds with modal operators and mood operators. The result will be a theory that only quantifies over particles. The language we are going to use for those translations is the relational modalese language $\mathcal{R}$. It extends standard quantified predicate logic with modal operators and mood operators, as described earlier. It also includes an identity predicate and a pair of non-logical predicates:

Babc $\quad$ *b* is between *ac*

Cabcd $\quad$ *a* is as far from *b* as *c* is from *d*

Notice that since $\mathcal{R}$ uses modal operators and mood operators instead of quantification over worlds, there is no need for anything like a world predicate, actuality predicate, or in predicate. The betweenness and congruence predicates are also no longer indexed to worlds.

$\mathcal{P}$ are $\mathcal{R}$ can both be naturally interpreted on modal ratio spaces.[20] Strictly speaking, these are just certain kinds of sets:

> A **modal ratio space** $\langle W, A, D, I, B, C\rangle$ consists of a non-empty W, a singleton $A \subset W$, a non-empty D, an $I \subset D \times W$, a $B \subset D \times D \times D \times W$, and a $C \subset D \times D \times W \times D \times D \times W$.

But while modal ratio spaces are just sets, it is generally more illuminating to think about them in terms of what they represent. That is, a modal ratio space consists of a domain of worlds W, precisely one of which has the property A of being actual. There is then a domain of particles D. Those particles are related by the in relation I to worlds. They are

20. Given that our ultimate goal is to do science without numbers, the appeal here to *sets* might seem a little strange. But for my own part, as a nominalist, I do not deny that sets exist. What I deny is that sets have the same sort of “robust” existence as particles. Particles are real features of the fundamental physical world. Sets and other abstracta are just convenient fictions or conventional constructions or something else along those lines. In the present case, the appeal to modal ratio spaces is warranted because they need not be part of the fundamental physical world. They are just a convenient device for translating sentences of possiblese into sentences of modalese, and visa-versa.

also related by the betweenness relation B and congruence relation C. The betweenness relation is indexed to a single world. The congruence relations is indexed to a pair of worlds. This means that while betweenness relations are always within worlds, congruence relations can hold across worlds.

The interpretation of $\mathcal{P}$ on modal ratio spaces is what you would expect. Names are assigned to elements of $W \cup D$. Variables range over $W \cup D$. The world predicate is assigned to W, the actuality predicate is assigned to A, the in predicate is assigned to I, the betweenness predicate is assigned to B, and the congruence predicates is assigned to C. The identity predicate is interpreted as genuine identity.

The interpretation of $\mathcal{R}$ is also straightforward. The interpretation function maps names to elements of D. Variable assignments map arbitrary terms to elements of D. In the case of names, they do so in a way that matches the interpretation function. We then recursively define satisfaction relative to a pair of worlds and a variable assignment.

$wv \models_\sigma B(t_1, t_2, t_3)$ iff $\langle \sigma(t_1), \sigma(t_2), \sigma(t_3), w \rangle \in B$

$wv \models_\sigma C(t_1, t_2, t_3, t_4)$ iff $\langle \sigma(t_1), \sigma(t_2), w, \sigma(t_3), \sigma(t_4), v \rangle \in C$

$wv \models_\sigma t_1 = t_2$ iff $\sigma(t_1) = \sigma(t_2)$

$wv \models_\sigma \exists x \phi(x)$ iff there is a τ such that $wv \models_\tau \phi(x)$

$wv \models_\sigma \neg(\phi)$ iff $wv \neg \models_\sigma \phi$

$wv \models_\sigma \phi \wedge \psi$ iff $w \models_\sigma \phi$ and $w \models_\sigma \psi$

$wv \models_\sigma \Diamond(\phi)$ iff $uw \models_\sigma \phi$ for some u

$wv \models_\sigma \otimes(\phi)$ iff $vw \models_\sigma \phi$

$wv \models_\sigma \downarrow(\phi)$ iff $vv \models_\sigma \phi$

The variable assignment τ in the clause for existential quantification is just like σ, with the possible exception that $\tau(x) \neq \sigma(x)$ and the additional requirement that $\langle \tau(x), w \rangle \in I$. A sentence ϕ is then true at a world w when $ww \models_\sigma \phi$ for all σ. A sentence ϕ is true under an interpretation when true at the actual world $@ \in A$.

Looking at the above semantics, you can see why relational modalese lets us solve the problem of expression. In standard modalese, satisfaction is defined relative to a single world. Basic predicates are then assigned to $n+1$ tuples, with the extra element used as a world index.[21] Relational modalese does things differently. It defines satisfaction relative to a pair of worlds. Basic predicates can then be assigned to $n+2$ tuples, with the two extra elements used as a pair of world indexes. This lets the language simulate relations across world. The role of mood operators, then, is to permute those pairs of worlds.[22]

Notice that we now also have a natural strategy for answering the question about

---

21. Or they are assigned to intensions, which are functions from worlds to sets of n tuples. These are equivalent to sets of $n+1$ tuples.

22. The idea of doing modal semantics with pairs of worlds is not completely new. It is fairly common to extend standard modalese with the Vlach operators $\downarrow$ and $\uparrow$, for example. Satisfaction is then defined relative to pairs of worlds. What is new is the idea that we can use this sort of semantics to express a genuinely *relational* notion of possibility, one that is reflected by our ordinary use of grammatical mood.

On the more standard Vlachean approach, basic n place predicates are still assigned to $n+1$ tuples of worlds, even though satisfaction is defined relative to pairs of worlds. Since we only have one world index to work with, there is no way to directly express relational possibilities. Another problem is that the set of basic operators is poorly chosen. If we start with wv, we can use $\downarrow$ to get to vv and $\uparrow$ to get to ww. The problem is that we have no way to get to vw. Vlach operators may let us permute pairs of worlds, then, but they fail to give us all the permutations.

worlds from the end of the last section. When is talk about possible worlds apt? A sentence of possiblese is apt, we might say, when it correctly translates a true sentences of relational modalese. And a sentence of possiblese correctly translates a sentences of relational modalese when the two sentences always have the same truth value whenever they are interpreted on the same modal ratio space.[23]

Now that we have both possiblese and relational modalese, we can set to work doing science. We can use possiblese to build a theory of distance ratios that quantifies over worlds and particles. Then we can use relational modalese to translate that theory into one that only quantifies over particles.

As you will recall, Tarski gives us a theory of distance ratios using betweenness and congruence relations between spacetime points. What we are going to call a **Tarski world** is a world in which his axioms for *points* are satisfied by *particles*. At Tarski worlds, then, there are no gaps between particles. There are no gaps between particles because particles have the structure of spacetime points. To say that a world is Tarski world in possiblese, we are going to use a defined predicate.

Tw particles have the structure of points at w

This predicate is defined by systematically translating Tarski's axioms, then letting the variables range over particles instead of points.[24]

---

23. And any names shared by the two languages are interpreted in the same way.

24. A bit more precisely, start with the conjunction ϕ of Tarski's axioms. This is a sentence in the language of his theory. We then replace his quantifiers with quantifiers restricted by the in predicate and replace his betweenness and congruence predicates with ours. We also make sure that all of the world indexes are filled

Once we have the Tarski predicate, we can complete the first stage of our project. We can give a theory of distance ratios using only possible worlds and particles. The theory has six axioms, which are listed below.[25]

P1 $\exists w(Ww \wedge \forall ab(Iabv \supset Cabwabv \wedge Tw)$

P2 $Cabwabw$

P3 $Cabwcdv \supset Ccdvabw$

P4 $Ccdvefu \wedge Cabwcdv \supset Cabwefu$

P5 $Cabwefv \wedge Cbcwfgv \wedge Cacwegv \wedge Babcw \supset Befgv$

P6 $Tw \wedge Tv \wedge Iabw \supset \exists cd(Icdv \wedge Cabwcdv)$

How does the theory work? Suppose the actual world v has exactly four particles *abcd*. The first axiom of our theory says that every world can be extended to a Tarski world. There is thus a possible Tarski world w at which *abcd* are as far apart as they are at v, but with all the gaps filled in by other particles. Now we know that distance ratios can be defined at w, since particles at w have the same structure as spacetime points. The particles *abcd* thus

with the same world variable w. The result is an open sentence $\phi(w)$ of possiblese that is satisfied by a world iff the particles in that world have the same structure as points. The predicate Tw is then satisfied iff the open sentence $\phi(w)$ is satisfied.

One wrinkle is Tarski's Archimedean axiom, which quantifies over sets of particles. Rather than using quantification over sets, we can use plural quantification over particles, but then we need a language with plural quantification. There is no problem adding plural quantification to possiblese, but it adds an extra layer of complexity. So we are ignoring plural quantification in the main text.

25. To simplify the notation, the wide scope universal quantifiers has been left implicit. Think of each axiom as being prefaced by a string of universal quantifiers $\forall w \forall v \forall a \forall b$... binding any free variables.

have a unique numeric distance ratio r at w. What we claim, then, is that because they are as far apart at w as they are at v, they *thereby* have the same numeric distance ratio at v. Particles in sparse worlds thus inherit their distance ratios from particles in Tarski worlds rather than from, say, a background spacetime.

That's the basic idea anyway. The remaining axioms then ensure that inherited distance ratios behave as expected. Axioms two, three, and four say that congruence across worlds is an equivalence relation. This makes distance ratios unique. Axiom five says that congruence preserves betweenness. This ensures that distance ratios sum correctly. Axiom six says that Tarski worlds are always comparable. This ensures that numeric distance ratios are always defined not only within worlds, but across worlds.[26] The result is a theory that fully fixes numeric distance ratios while quantifying over nothing but worlds and particles.[27,28]

---

26. That is, the theory ensures that for any particles ab in any world w and any particles cd in any world v, there is unique numeric distance ratio r of ab at w to cd at v. Whether these sorts of distance comparisons across worlds are a vice or a virtue is an open question. Dasgupta (2014) argues that they are a vice. Baker (2015) gives reason to think they are virtue. If you think they are a vice, then we could avoid such comparisons by using an eight place distance ratio congruence predicate in place of the four place distance congruence predicate.

27. Similar proposals can be found in (Manders 1982) and (Belot 2011). Manders quantifies over what he calls configurations which, unlike worlds, can be parts of other configurations. Belot quantifies over things like metric spaces, sets, and isometries, so his theory is no help for doing science with only particles.

28. There are various ways in which the above theory could be improved. For example, you might wonder if Tarski worlds are really possible. But in fact, this is just a simplifying assumption. Rather than using Tarski worlds, we could go through Tarski's axioms one by one and find appropriate substitutes. Done properly, this only requires worlds with finitely many particles, so there is no need for the possibility of Tarski worlds.

We thus have a theory of distance ratios in possiblese. The final step is to translate that theory into relational modalese. The first axiom, for example, can be translated as:

M1 $\diamond(\otimes\forall ab\otimes(Cabab \wedge \uparrow T))$

The first swap operator puts the universal quantifier into the indicative mood and the second puts everything else back into the default mixed mood. So $\diamond(\otimes\forall ab\forall Cabab \wedge ...)$ says that it could have been that for any two particles there are, those particles were as far apart as they are and... The up operator then puts T uniformly into the indicative mood. Here, T is just the conjunction of Tarski's axioms, with the variables understood as ranging over particles. We can read $\diamond(...\uparrow T)$, then, as saying that it could have that...particles had the same structure as points. Putting these two pieces together:

> It could have been that for any particles *ab* there are, *ab* were as far apart as *ab* are while particles had the same structure as points.

We thus have an intuitively correct translation of (P1). The translation is also correct in the more precise sense that that (P1) and (M1) have the same truth value whenever they are interpreted on the same modal ratio space.

So far so good. The problem is that while relational modalese can translate *most* of the remaining axioms, it cannot translate them all. The problematic axiom is number four, which says that the congruence across worlds is transitive.

P4 $Cabwcdv \wedge Ccdvefu \supset Cabwefu$

The problem, basically, is that relational modalese is a language for describing possibilities taken two at a time. But crossworld transitivity principles require talking about possibilities taken *three* at a time. In terms of possible worlds, we need to say that for any three worlds w, v, and u if particles from the first are congruent with particles from the second and particle from the second are congruent with particles from the third, then particles from the first are congruent with particles from the third. We thus need to describe three worlds at once, even though the congruence relation can only span two worlds at once.

The solution is to use a slightly more powerful version of relational modalese. Standard modalese can describe possibilities one at a time. The original relational modalese language can describe possibilities taken two at a time, so call that language $\mathcal{R}_2$. The more powerful language we need can describe three possibilities at a time. That language is $\mathcal{R}_3$.

$\mathcal{R}_3$ extends the language of standard quantified predicate logic with the same operators as $\mathcal{R}_2$. The difference is that it has two swap operators $\otimes_2$ and $\otimes_3$ instead of just one. The interpretation of $\mathcal{R}_3$ on a modal ratio spaces is then basically the same as before. Names are assigned to elements of D. Variables range over the elements of D. The betweenness predicate is assigned to B and the congruence predicate is assigned to C. The difference is that satisfaction is defined relative to *triples* of worlds instead of pairs of worlds:

$wvu \models_\sigma B(t_1, t_2, t_3)$ iff $\langle \sigma(t_1), \sigma(t_2), \sigma(t_3), w\rangle \in B$

$wvu \models_\sigma C(t_1, t_2, t_3, t_4)$ iff $\langle \sigma(t_1), \sigma(t_2), w, \sigma(t_3), \sigma(t_4), v\rangle \in C$

$wvu \models_\sigma t_1 = t_2$ iff $\sigma(t_1) = \sigma(t_2)$

$wvu \models_\sigma \exists x \phi(x)$ iff there is a τ such that $wvu \models_\tau \phi(x)$

$wvu \models_\sigma \neg\phi$ iff $wvu \neg \models_\sigma \phi$

$wvu \models_\sigma \phi \wedge \psi$ iff $wvu \models_\sigma \phi$ and $wvu \models_\sigma \psi$

$wvu \models_\sigma \Diamond(\phi)$ iff $zwv \models_\sigma \phi$ for some z

$wvu \models_\sigma \otimes_2(\phi)$ iff $vwu \models_\sigma \phi$

$wvu \models_\sigma \otimes_3(\phi)$ iff $wuv \models_\sigma \phi$

$wvu \models_\sigma \downarrow(\phi)$ iff $vvu \models_\sigma \phi$

As you can see, the $\otimes_2$ operator exchanges the first world with the second. And the $\otimes_3$ operator exchanges the first with the third.

Once we have the two swap operators, we can translate all six axioms, including the recalcitrant transitivity axiom.[29]

M1 $\Diamond\otimes_2\forall ab\otimes_2(Cabab \wedge \uparrow T)$

M2 $\uparrow Cabab$

M3 $Cabcd \supset \otimes_2 Ccdab$

M4 $Cabcd \wedge \otimes_3\otimes_2 Ccdef \supset \otimes_3\otimes_2\otimes_3 Cabef$

M5 $Cabef \wedge Cbcfg \wedge Caceg \wedge \uparrow(Babc) \supset \downarrow(Befg)$

M6 $T \wedge \downarrow T \wedge \forall ab\, \otimes_2\exists cd\otimes_2\, Cabcd$

As before, each axiom is a correct translation of the original axiom in the sense that the two sentences are guaranteed to have the same truth value when interpreted on the same modal ratio space. This can be easily verified. The result is a theory that works the same way as the original and fixes numeric distance ratios for the same reason. The difference is that now, our quantifiers only range over particles.[30]

29. To keep the notation compact, we are suppressing the wide scope necessity operators and wide scope universal quantifiers. Think of each axiom as prefaced by a string of three boxes □□□... followed by a string of universal quantifiers $\forall a \forall b$... binding any free variables.

30. This basic strategy can be extended to explain the Minkowski spacetime relations needed to do special

Now, there is an interesting question about the extent to which $\mathcal{R}_3$ reflects our ordinary modal talk. Myself, I suspect that $\mathcal{R}_2$ is a better match. Ordinary language would seem to have *two* grammatical moods—the indicative and the subjunctive—and so ordinary modal auxiliaries would seem to be built for describing possibilities taken two at a time.

That said, there are sentence of natural language that arguably describe three possibilities at once. For example:

> Had Socrates been taller than he is, then had Socrates been taller than *that*, he would have been taller than he is.

Thinking of counterfactuals as strict conditions, the natural reading of this sentence is one on which it is true iff for all possible worlds wvu, if Socrates at v is taller than Socrates at u and Socrates at w is taller than Socrates at v, the Socrates at w is taller than Socrates at u. The question then is about how to understand the anaphoric reference. Does 'that' anaphorically refer to a *height*? Or does it anaphorically refers Socrates *as he were*? I'm not entirely sure myself.

---

relativity. Mundy (1986) shows how to axiomatize Minkowski spacetime using a trio of betweenness relations (lightlike betweenness, spacelike betweenness, and timelike betweenness) and a pair of congruence relations (spatial congruence and temporal congruence). Replacing his points with particles and indexing his predicates to possible worlds, we can give a possiblist theory using what we might call Mundy worlds in place of Tarski worlds. Once we have that, we can translate the possiblist theory into relational modalese, which eliminates the quantification over possible worlds. We thus have a modal theory of Minkowski spacetime relations that quantifies over nothing but particles. The interesting question then is whether this basic approach can be extended to Riemannian spacetime relation—the sort of spacetime relations needed for general relativity. That is an open question that deserves further research.

Even if English has no native mechanism for making such claims, the fact that English has only two grammatical moods would not seem to be an especially deep feature. There is a close analogy between mood and tense and, when it comes to tense, we can distinguish between not just the past and present, say, but the past, present, and *past perfect*. We can say, for example, that

Socrates was taller than he *had been*, even though he was not taller than he is.

This lets us make claims involving three times at once rather than just two. But if tense can work like that, then there is no obvious reason that *mood* could not work like that. We could have easily found ourselves speaking a language that could distinguish between not just the indicative and subjunctive, but the indicative, subjunctive, and the *perfect* subjunctive. And in that case, we would be speaking a natural language that was a better match for $\mathcal{R}_3$.

Fortunately, however things shake out with respect to natural language, none of this is a serious barrier to using $\mathcal{R}_3$ for science. When doing science, we always start with our ordinary concepts. We start with our ordinary ways of thinking about things like space and time and matter. We often find, though, that those ordinary concepts need to be somewhat reshaped and revised. It would not be surprising, then, if science required somewhat revising our ordinary modal concepts. Our modal concepts—like all the others—can be reshaped by reflecting on the scientific project. And so it might be that $\mathcal{R}_2$ best reflects our ordinary modal thought, but that $\mathcal{R}_3$ is the right language for describing the fundamental physical world.

1.7. Conclusion

Relational possibilities would seem to require quantification over things like numbers, distances, and spacetime points. But if so, then we cannot very well use relational possibilities to do science with only particles, since we will have to quantify over precisely the sort of things we are trying to eliminate.

As we have seen, though, the apparent need to quantify over further things is really just an artifact of a certain *way* of expressing relational possibilities. Standard modalese requires quantification over further things because it has modal operators, which correspond to modal auxiliaries, but nothing corresponding to grammatical mood. But once we *have* a language that has both modal operators and mood operators—once we have relational modalese—we can express relational possibilities directly. There is no need for quantification over further things. This not only better reflects what we do in natural language. It give us a powerful strategy for doing science with only particles.

Now in section six, we built a relational modalese language to solve a particular problem—we wanted to give an axiomatic theory of distance ratios. For those particular purposes, it proved useful to interpret the language on what we called modal ratios spaces. For relational modalese to have more general application, though, we also need more general tools. For those interested, more on that can be found in the appendix.

Relational modalese is not just a tool for doing science with only particles. It represents a new way of thinking about the nature of modality. Modality is about how things could have *differed*, not just how things could have been. With some luck, this new perspective will help us solves other sorts of puzzles as well.

2 The Paradox of Counterfactual Tolerance

Counterfactuals are somewhat tolerant. Had Socrates been at least six feet tall, he need not have been exactly six feet tall. He might have been a little taller—he might have been six one or six two. But while he might have been a little taller, there are also limits to how tall he would have been. He would not have been a thousand feet tall, for example. Counterfactuals are not just tolerant, then, but bounded.

This paper presents a surprising paradox. If counterfactuals are both tolerant and bounded, then we can prove a flat contradiction using natural rules of inference. Moreover, not only are these rules of inference natural, they are also generally validated by our best semantic theories. These include the familiar Lewisian analysis in terms of similarity. Something has to go then. But what?

Putting my own cards on the table: I think that counterfactuals are both tolerant and bounded, and that this places an important constraint on counterfactual semantics. My own solution to the problem is to deny that similarity is transitive. This corresponds to analyzing counterfactuals in terms of *sufficient* similarity instead of precise similarity. We will of course have much more to say about all this as we go along. But first, let's start with the paradox.

2.1. Paradox

Planck lengths are incredibly small. You would quite literally need a hundred million trillion of them just to span the diameter of a proton. Using Planck lengths as our basic unit of measure, we claim that counterfactuals are both tolerant and bounded.

> **Tolerance:** For any h, had Socrates been at least h, he might have been at least $h + 1$.
>
> **Boundedness:** There are j and k such that had Socrates been at least j, he would not have been at least k.

Tolerance says that had Socrates been at least six feet, he might have been at least a Planck length taller, and likewise for other heights. Boundedness says that there are heights j and k such that k is a **counterfactual bound** of j. So for example, it might be that had Socrates been at least six feet, he would not have been at least a thousand feet. In that case, a thousand feet will be a counterfactual bound of six feet, and boundedness will be satisfied.

Besides thinking that counterfactuals are tolerant and bounded, we also think that certain inferences preserve truth. For example, suppose that had the Athenians invaded Sparta, they would have lost and might have used catapults. It would then seem to follow that had they invaded Sparta *and* used catapults, they would have lost. This reasoning is perfectly natural and backed by an axiom called **rational monotonicity**.

$$\text{RM} \quad (A \mathrel{\Box\!\!\rightarrow} C) \wedge (A \mathrel{\Diamond\!\!\rightarrow} B) \supset (A \wedge B \mathrel{\Box\!\!\rightarrow} C)$$

As you can see, rational monotonicity is a kind of restricted strengthening rule. It tells us that we can strengthen from $A \mathrel{\Box\!\!\rightarrow} C$ to $A \wedge B \mathrel{\Box\!\!\rightarrow} C$ under certain special circumstances.

Those special circumstances are that $A \mathbin{\Diamond\!\!\rightarrow} B$.[1]

We can now prove a flat contradiction. Below, h represents an open sentence saying that Socrates is at least h Planck lengths tall. Likewise for k and $n+1$ and so on. Classical predicate logic is used throughout.

(1)	$j \mathbin{\Box\!\!\rightarrow} \neg k$	boundedness
(2)	$n \mathbin{\Box\!\!\rightarrow} \neg k$	hypothesis
(3)	$n \mathbin{\Diamond\!\!\rightarrow} n+1$	tolerance
(4)	$n \wedge (n+1) \mathbin{\Box\!\!\rightarrow} \neg k$	2, 3, RM
(5)	$n+1 \mathbin{\Box\!\!\rightarrow} \neg k$	4, substitution
(6)	$k-1 \mathbin{\Box\!\!\rightarrow} \neg k$	1, 2, 5, induction
(7)	$k-1 \mathbin{\Diamond\!\!\rightarrow} k$	tolerance
(8)	$\bot$	6, 7, duality

We have proved absurdity, so something has to go. Either one of our premises is false or one of our inference rules is invalid. Call this the **tolerance paradox**.

2.2. Premises

Our paradoxical argument has two premises and, so, we might try to resolve the paradox by denying one or the other.

1. As we go along, we will sometimes run together talk about *rules* and *conditional axioms*. The reason being that for present purposes, nothing much hangs on this distinction, and keeping track of it would be tedious. But we should point out that officially, RM is a conditional axiom. Given modus ponens and conjunction introduction, this makes it strictly stronger than the corresponding rule $A \mathbin{\Box\!\!\rightarrow} C, A \mathbin{\Diamond\!\!\rightarrow} B \vdash A \wedge B \mathbin{\Box\!\!\rightarrow} C$.

Suppose we deny boundedness. This means accepting that had Socrates been any taller, he might have been any height whatsoever. But any height whatsoever? We generally think that even if Socrates had been taller, the Spartans would not have thrown their weapons into the sea. But if Socrates might have been any height *whatsoever*, this is simply false. He might have been tall enough to crush Sparta with a single step. And in that case, the Spartans *would* have thrown their weapons into the sea. Who wants to oppose the Great Giant of Athens? If we deny that counterfactuals are bounded, so many ordinary counterfactuals will turn out false that we might as well give up counterfactual reasoning altogether.

Two more things should be pointed out about boundedness. First, boundedness says that some height has a counterfactual bound. It does not say that *all* heights have a counterfactual bound. Sufficiently outlandish heights might still be unbounded. For all boundedness cares, it may be that had Socrates been at least a hundred million trillion feet tall, he might have been any height whatsoever. Even so, boundedness will still hold if there is some pair of heights j and k such that had Socrates been at least j, he would not have been at least k. Six feet and a thousand feet would seem to do the trick. Even if Socrates might have been any height whatsoever, had he been at least a hundred million trillion feet, he would not have been at least a thousand feet, had he been at least six feet.

Second, boundedness does not require there to be a height with a *least* counterfactual bound. If it did, there might be concerns about sharp cutoffs and whether or not boundedness was determinately true. But since boundedness only requires there to be a height with some counterfactual bound, those concerns can be sidestepped. It may be vague whether or not Socrates might have been at least twelve feet, had he been at least six feet. But even if so, it is *not* vague whether or not he might have been a thousand feet.

He determinately would not have been.

Suppose that instead of denying boundedness, we deny tolerance. This means accepting the existence of what we might call **singularities**. Singularities are just heights such that had Socrates been at least that tall, he would not have been even a Planck length taller.

As it turns out, there are certain good reasons to accept the existence of singularities. One of them is **strong centering**.

SC $(A \wedge B) \supset A \mathrel{\Box\!\!\rightarrow} B$

This says that whenever A and B are both true, then $A \mathrel{\Box\!\!\rightarrow} B$ is also true. So for example, say that Socrates is exactly five feet tall. It then follows that he is at least five feet and that he is not at least a Planck length taller. So by strong centering, had Socrates been at least five feet, he would not have been even a Planck length taller. The actual height of Socrates thus turns out to be a singularity, given strong centering.

Another reason to think that there are singularities is that certain heights may be impossible. For example, it may be biologically possible for Socrates to have been up to fifteen feet tall, but biologically impossible for him to have been any taller. In that case, fifteen feet would seem to be a singularity, and so a counterexample to tolerance.

Why is that? There are certain plausible connections between possibility, necessity, and counterfactuals. For example, it would seem that for any genuine notion of possibility and corresponding notion of necessity, a would counterfactual is true whenever the antecedent is possible and the consequent is necessary. That is:

M $(\Diamond A \wedge \Box B) \supset A \mathrel{\Box\!\!\rightarrow} B$

This rule correctly predicts that since Socrates is possibly a farmer and necessarily human, it follows that had Socrates been a farmer, he would have been human.[2] This is exactly the sort of thing you would expect.

Now suppose that biological possibility is a genuine form of possibility. And suppose that it is biologically possible for Socrates to have been fifteen feet tall, but biologically impossible for him to have been any taller. This means that it is biologically possible for Socrates to have been at least fifteen feet and biologically necessary that he is not at least fifteen feet plus a Planck length. Given the above inference rule, it follows that had Socrates been at least fifteen feet, he would not have been at least fifteen feet plus a Planck length. So fifteen feet is a singularity and an apparent counterexample to tolerance.

Special circumstances can also lead to singularities. Suppose that Socrates is in fact genetically engineered by a team of alien scientists. Their advanced technology lets them precisely control his height and, as it turns out, their research grant requires them to make his height a prime number of Planck lengths. But research funding being what it is, the scientists also have a strong preference for smaller heights, since smaller humans are cheaper to build. The result is that for every prime height greater than his actual height, had Socrates been at least that tall, he would have been exactly that tall. All those heights are singularities.

There are different ways of dealing with such difficulties. We could deny strong centering. We could deny the proposed link between possibility, necessity, and would counterfactuals. The simplest solution, though, is to just restrict tolerance and boundedness.

2. This is similar to a rule discussed at some length in (Lange 2009, pp. 64-67).

What we claim is that there is some *range* of heights R such that for every height $h \in R$, had Socrates been at least h, he might have been at least $h+1$. And for some pair of heights $j, k \in R$, had Socrates been at least j, he would not have been at least k. This is all we need to run the paradox. The original way we formulated tolerance and boundedness, while simpler, is somewhat stronger than what we actually need.

That there are such ranges of heights would seem to be obvious. For example, take the range of heights between six feet and nine feet. Had Socrates been at least six feet, he would not have been at least nine feet. And for every height between six feet and nine feet, had Socrates been at least that tall, he might have been a Planck length taller. Strong centering is no longer a problem, since Socrates is actually five feet, and five feet is well below the bottom of the range. And the impossibility of Socrates being taller than fifteen feet is no longer a problem, since fifteen feet is well above the top of the range.

Putting the issue in the most general terms: The rules of counterfactual inference are necessarily valid, if valid at all. So all we need is for it to be *possible* for there to be *some* range of *some* quantity that satisfies both tolerance and boundedness. If so, then we can reason to paradox.

2.3. Auxiliary Inferences

The tolerance paradox uses various auxiliary inference rules and, so, you might wonder if they are the source of the problem. These include classical predicate logic and mathematical induction, which are not promising points of resistance. Giving up either means giving up modern mathematics, and we are not going to give up modern mathematics. The cost is simply too high. Moreover, even if we *were* willing to give up induction or classical logic, it is not especially clear how that would help. Our paradoxical proof is intuitionistically valid, so valid by the lights of the most promising alternative to classical logic. And while induction is convenient, it is also not essential. Suppose that we agree that some particular height has a certain counterfactual bound. Suppose we agree that had Socrates been at least six feet, he would not have been at least nine feet. We could then show that had Socrates been at least a thousand feet, he would not have been even a Planck length taller using a very long—but still finite—proof. The proof would use the same basic reasoning, but not require induction.

This leaves two remaining auxiliary rules. The first is **substitution**.[3] Substitution says that when two propositions are equivalent, we can replace the one with the other in the antecedent of a counterfactual. Equivalent in what sense? There are several options here. My own inclination is to use counterfactual equivalence.[4] But the relevant propositions are also necessarily equivalent, a priori equivalent, analytically equivalent, and maybe even logically equivalent. Basically, *any* reasonable rule telling us that we can substitute equivalents will let us replace the proposition that Socrates is at least n and at least $n+1$ with the proposition that he is at least $n+1$.

Moreover, even if all such substitution principles were to fail, the inference would still go through by other means. Thinking of substitution as substitution under counterfactual equivalence and making the counterfactual equivalence explicit, the original reasoning

---

3. $A \mathrel{\Box\!\!\rightarrow} C \supset B \mathrel{\Box\!\!\rightarrow} C$ when A and B are equivalent.

4. A and B are counterfactually equivalent when $A \mathrel{\Box\!\!\rightarrow} B$ and $B \mathrel{\Box\!\!\rightarrow} A$.

looks like this:

(4)	$n \wedge (n+1) \boxright \neg k$	2, 3, rational monotonicity
(5)	$(n \wedge (n+1) \boxright n) \wedge (n \boxright n \wedge (n+1))$	nature of heights
(6)	$n+1 \boxright \neg k$	4,5, substitution

But even without substitution, we can get the same effect using other rules. For example, we could use **limited transitivity** to reason as follows.[5]

(4)	$n \wedge (n+1) \boxright \neg k$	2, 3, rational monotonicity
(5)	$n+1 \boxright n$	nature of heights
(6)	$n+1 \boxright \neg k$	4,5, limited transitivity

The rest of the proof then goes through as before. This means that denying substitution—on its own at least—will not resolve the paradox.

The other auxiliary inference rule is **duality**.[6] Duality says there is a certain equivalence between might and would counterfactuals. This is widely accepted, but also controversial. Fortunately, while posing the paradox using duality is natural, it is also not required. One

5. Limited transitivity tells us is that whenever we have a would counterfactual with conjunctive antecedent and one of the conjuncts counterfactually entails the other, the conjunct that is counterfactually entailed can be eliminated. In the present case, it is probably best thought of as a pair of axioms:

$$(A \boxright B) \wedge (A \wedge B \boxright C) \supset A \boxright C$$

$$(B \boxright A) \wedge (A \wedge B \boxright C) \supset B \boxright C$$

The reason that we need two axioms is that we need one for each conjunct. Of course, if we had *substitution*, then only the first would be needed, since the antecedents are logically equivalent. But the status of substitution is currently in question, so we need two.

6. $A \boxright B \equiv \neg(A \diamondright \neg B)$

reason is that in the original argument, we used induction and duality to reason as follows:

(6) $k-1 \;\Box\!\!\rightarrow \neg k$ — 1, 2, 5, induction

(7) $k-1 \;\Diamond\!\!\rightarrow k$ — tolerance

(8) $\bot$ — 6, 7, duality

But even without duality, we could still have used induction to directly inferred that:

(6) $k \;\Box\!\!\rightarrow \neg k$ — 1, 2, 5, induction

This is bad enough. Thinking of k as a thousand feet, this says that had Socrates been at least a thousand feet, he would *not* have been at least a thousand feet. This strikes me as not just false, but inconsistent.[7] So we still have paradox. Denying duality is no help.[8]

2.4. Rational Monotonicity

Rational monotonicity is natural and part of counterfactual common sense. Had the Athenians invaded Sparta, they would have lost and might have used catapults. What could be more natural than concluding that had they invaded *and* used catapults, they would have lost?

---

7. Why think that $\vdash \neg(k \;\Box\!\!\rightarrow \neg k)$? First, we claim that $\vdash (\Diamond A) \wedge (A \;\Box\!\!\rightarrow B) \supset \Diamond(A \wedge B)$ for any genuine notion of possibility and necessity. But now take the case in which $\Diamond$ and $\Box$ denote *logical* possibility and necessity respectively. Since $\nvdash \neg k$ and $\vdash \neg(k \wedge \neg k)$, we thus have $\vdash \Diamond k$ and $\vdash \neg\Diamond(k \wedge \neg k)$. But then since logical possibility and necessity are genuine forms of possibility and necessity, it follows that $\vdash \neg(k \;\Box\!\!\rightarrow \neg k)$ by the proposed principle and propositional logic.

8. The other reason that denying duality is no help is that we can reformulate the original paradox using only might counterfactuals. That version of the argument can be found in section 2.11.

Besides being natural, rational monotonicity is also generally validated by our best counterfactual semantics. A prominent example is the similarity semantics from **Lewis1973**. He analyzes counterfactuals using what we are going to call **precise similarity models**. These are triples:

$$\langle W, R, [\![\cdot]\!] \rangle$$

where W is a set of worlds v is a valuation function. R is a three-place counterfactual accessibility relation on worlds. This means that, strictly speaking, R is just a set of triples.[9] But Lewis think of it as a *similarity* relation on worlds, so we can too.

$Rwvu$ w is no less similar than v to u

Because Lewis thinks of R as precise similarity, he requires it to be total and transitive and so form a weak total ordering.

Total	either $Rwvu$ or $Rvwu$
Transitive	if $Rwvz$ and $Rvuz$, then $Rwuz$

It would seem to be clear that the similarity relation has these features. Suppose we are comparing similarity relative to some fixed world z. Totality then says that there are never pairs of worlds such that each is less similar than the other. Transitivity says that if w is no less similar than v and v is no less similar than u, then w is no less similar than u.

Once we have our strict similarity models, we can define the counterfactual operators.

9. That is, $R \subset W^3$.

$u \models A \mathrel{\Box\!\!\rightarrow} B$ iff there is a world $v \in W$ such that $v \models A$ and every world $w \in W$ is such that $w \models A \supset B$ whenever $Rwvu$

$u \models A \mathrel{\Diamond\!\!\rightarrow} B$ iff for every world $v \in W$ such that $v \models A$, there is a world $w \in W$ is such that $w \models A \wedge \neg B$ and $Rwvu$

To simplify matters a bit, suppose we have the **limit assumption** in place.[10] What these clauses are then telling us is that:

> $A \mathrel{\Box\!\!\rightarrow} B$ is true iff all of the most similar A worlds are B worlds.
>
> $A \mathrel{\Diamond\!\!\rightarrow} B$ is true iff some of the most similar A worlds are B worlds.

Here, an A world is among the most similar A worlds when it is no less similar than any other A world.

But now here is something we can prove. We can prove that rational monotonicity is valid over the class of all strict similarity models.[11] This means that if we are going to

---

10. The limit assumptions says that there are no infinite descending chains of increasing similarity. More formally, for every $X \supset W$ and any $u \in W$, there is a $w \in X$ such that $Rwvu$ for every $v \in X$.

11. We need to show that $\models (A \mathrel{\Box\!\!\rightarrow} C) \wedge (A \mathrel{\Diamond\!\!\rightarrow} B) \supset (A \wedge B \mathrel{\Box\!\!\rightarrow} C)$. Suppose then that $A \mathrel{\Box\!\!\rightarrow} C$ and $A \mathrel{\Diamond\!\!\rightarrow} B$ are both true at a world z. Given the semantics, this means that there is a world v at which A such that all of the worlds w that are at least as similar to z are worlds at which $A \supset C$ and that one of those worlds u is a world at which $A \wedge B$. Now to show that $A \wedge B \mathrel{\Box\!\!\rightarrow} C$ is true at z, all we need to show is that every world u^* that is at least as close to z as u is a world at which $A \wedge B \supset C$. To show that, suppose otherwise. Suppose there is a world u^* that is at least as close to z as u and also a world at which $A \wedge B \wedge \neg C$. Then since worlds are logically closed, u^* is a world at which $A \wedge \neg C$. But then by transitivity, since u^* is at least as close as u to z and u is at least as close as v to z, that u^* is at least as close as v to z. But then $A \mathrel{\Box\!\!\rightarrow} C$ is false, contrary to assumption. So $A \wedge B \mathrel{\Box\!\!\rightarrow} C$ is true at z. Therefore $\models (A \mathrel{\Box\!\!\rightarrow} C) \wedge (A \mathrel{\Diamond\!\!\rightarrow} B) \supset (A \wedge B \mathrel{\Box\!\!\rightarrow} C)$.

resolve the counterfactual tolerance paradox by denying rational monotonicity, we will have to reject the Lewisian analysis of counterfactual in terms of similarity.

2.5. Uniform Intolerance

Now to my knowledge, Lewis has no official position on the tolerance paradox, but his earlier response to Pollock is telling. Lewis would rather give up infinite agglomeration than give up the idea that for every length longer than an inch, had the line been longer, it would have been shorter than that. This suggests that not only would Lewis deny that counterfactuals are tolerant, he would do so with characteristic gusto. He would claim that counterfactuals are **uniformly intolerant**. Letting g be the actual height of Socrates, Lewis would say that for every $h > g$, had Socrates been at least h, he would have been exactly h. This because any world in which Socrates is taller than h is less similar to our world than a world in which he is exactly h. It's not just that some of the heights greater than g are singularities. All of them are. This follows from the analysis of counterfactuals in terms of similarity.

There are certain advantages to uniform intolerance. In comparison to coarse graining, it removes any need to explain why some heights are singularities and others are not. But while that may be, uniform intolerance runs roughshod over how we ordinarily think about counterfactuals.

Suppose for example that Nixon has a finicky nuclear button. To prevent accidental launches, the button has to be pressed at least n Planck lengths to send a signal. But even when pressed at least $n + m$ Planck lengths, there is no guarantee that it launches a rocket.

If m is even, the button sends a weak signal that fizzles out before reaching the launch pad. If m is odd, the button sends a strong signal that launches a rocket. As it happens, Nixon presses the button, but not quite hard enough to send a signal. What would have happened if Nixon had pressed a little harder? Uniform intolerance tells us that everything would have been fine. Had Nixon pressed the button at least n Planck lengths, he would have pressed it exactly n Planck lengths. A weak signal would have been sent and it would have fizzled out. But this is the wrong result: Had Nixon pressed hard enough to send a signal, he might have pressed it a little past n. He might have launched a rocket and might have started a war.

We know that Nixon might have launched a rocket, in part, because of the role counterfactuals play in justifying our emotions. We should be *relieved* that Nixon did not push the button hard enough to send a signal. But if we know that had Nixon pushed the button, he would not have launched a rocket, why are we relieved? Nothing bad would have happened. Or suppose that Kissinger offers Nixon a dollar to push the button lightly and two dollars to push it hard enough to send a signal. Nixon pushes it lightly to get the dollar, but then reflects: Had he pressed the button hard enough to send a signal, he would not have a launched a rocket, but would have gotten an extra dollar. He thus *regrets* not pushing the button. All he did was leave money on the table. But this is obviously silly. Nixon should have no regrets because, had he pushed the button hard enough, he would have gotten an extra dollar, but he also might have started a nuclear war. He did the right thing.

Putting the issue more generally: Lewis has built his semantics to avoid certain kinds of arbitrary choices. Had you flipped a fair coin, it might have landed heads and it might have landed tails. It is simply false that it *would* have landed heads. If it were true that it would have landed heads, then reality would have to somehow arbitrarily choose the heads scenario over the tails scenario, despite the fact that those scenarios are symmetric and so equally similar.[12] The problem with uniform intolerance is that reality will have to make choices that are *almost* completely arbitrarily. Perhaps the world where Nixon pushes the button n Planck lengths is marginally more similar to our world than the one where he pushes it $n+1$ Planck lengths. But the choice between these worlds is still *mostly* arbitrary. The choice between them is just about as unmotivated as the choice between the world where you flip heads and the world where you flip tails. You would think, then, that this is the sort of choice Lewis would want to avoid.

2.6. Vagueness

At this point, you might be convinced that we have a paradox, but concerned that we do not have a new paradox. Why is this not just the sorites paradox in another form?

To build a sorites paradox, we need a scale and something like a vague all-or-nothing predicate. We might observe, for example, that one grain of sand is not a heap, but that a thousand grains of sand is a heap. The scale is the number of grains of sand. The vague predicate is being a heap. We then reason as follows:

12. Stalnaker of course has a line of response to this sort of reasoning. We will say more about his selection models in section 2.11

(1)	One grain is not a heap.	premise
(2)	A thousand grains are a heap.	premise
(3)	It is not the case that (n grains are not a heap and $n+1$ grains are a heap).	premise
(4)	If n grains are not a heap, then $n+1$ grains are not a heap.	3, PL
(5)	A thousand grains are not a heap.	1,4, induction
(6)	$\bot$	2,5, PL

As you can see, the paradox gets going because we deny that there are any sharp cutoffs. We deny that there is any number of grains n such that n is not a heap, but that $n+1$ is a heap. But then given the denial of sharp cutoffs, we can reason to a flat contradiction using induction and classical logic.

Now in the case of counterfactuals, we certainly *can* build a sorites paradox. We can use heights as the scale and the counterfactual operators in place of a vague all-or-nothing predicate. Like the original sorites argument, that argument depends on the denial of sharp cutoffs. What we deny in the case of counterfactuals is that there is any sharp cutoff in how tall Socrates might have been.

> **No Sharp Cutoffs:** There are no heights m and n such that (a) it is true that had Socrates been at least m, he might have been at least n and (b) it is false that had Socrates been at least m, he might have been at least $n+1$.

Besides denying that there are sharp cutoffs, we also have a boundedness condition. We claim that:

Boundedness: There are heights m and k such that $m < k$ and it is not the case that had Socrates been at least m, he might have been at least k.

The sorites paradox for counterfactuals then runs as follows:

(1)	$m \diamond\!\!\rightarrow m$	theorem
(2)	$\neg(m \diamond\!\!\rightarrow k)$	boundedness
(3)	$\neg(m \diamond\!\!\rightarrow n \wedge \neg(m \diamond\!\!\rightarrow n+1))$	no sharp cutoffs
(4)	$(m \diamond\!\!\rightarrow n) \supset (m \diamond\!\!\rightarrow n+1)$	3, PL
(5)	$m \diamond\!\!\rightarrow k$	1,4 induction
(6)	$\bot$	2,5, PL

Think of m as six feet and k as a thousand feet. The paradoxical argument thus starts with the observation that had Socrates been at least six feet, he might have been at least six feet. This will be a theorem of any reasonable counterfactual logic. We then deny that had Socrates been at least six feet, he might have been at least a thousand feet and deny that there is a sharp cutoffs with respect to how tall Socrates might have been, had he been at least six feet tall. But then by classical logic and induction, we get a flat contradiction.

The key observation is that while the sorites paradox is certainly a paradox, the sorites paradox is not the tolerance paradox. The two paradoxes are distinct. They have different premises and rely on different inference rules.

Start with the premises. Both paradoxes rely on a kind of boundedness condition, so they have that much in common. The difference is that where the tolerance paradox depends on the acceptance of tolerance, the sorites paradox depends on the denial of sharp cutoffs. These are clearly different claims. On the one hand, the denial of tolerance entails the acceptance of sharp cutoffs. After all, if we deny tolerance, then there is some m such that:

$$\neg(m \Diamond\!\!\rightarrow (m+1))$$

But from this it follows that:

$$m \Diamond\!\!\rightarrow m \wedge \neg(m \Diamond\!\!\rightarrow (m+1))$$

So there is some n such that:

$$m \Diamond\!\!\rightarrow n \wedge \neg(m \Diamond\!\!\rightarrow \neg(n+1))$$

That is, there is a sharp cutoff with respect to how tall Socrates might have been, had he been at least m tall. On the other hand, going the other direction, the acceptance of tolerance *does not* entail the denial of sharp cutoffs. You might, for example, accept that for every m:

$$(m \Diamond\!\!\rightarrow m+1) \wedge \neg(m \Diamond\!\!\rightarrow m+2)$$

Had Socrates been at least six feet tall, he might have been at least one Planck length taller, but would not have been at least two Planck lengths taller, and likewise for other heights. But in that case, we have both tolerance and sharp cutoffs. The acceptance of tolerance follows from the first conjunct. Letting $n = m+1$, the acceptance of sharp cutoffs also follows, since:

$$(m \Diamond\!\!\rightarrow n) \wedge \neg(m \Diamond\!\!\rightarrow n+1)$$

In that case, the tolerance paradox will go through, but the sorites paradox will not. So the two paradoxes are not the same. The acceptance of tolerance is *strictly weaker* than the denial of sharp cutoffs.

Another reason to think that the paradoxes are distinct is that they rely on different inference rules. If you look back at the sorites paradox for counterfactuals, we used counterfactual logic to get the theorem on the first line, which was the claim that had Socrates been at least m, he might have been at least m. But we could have just taken that as a premise and, in that case, we would have had a paradox without using *any* counterfactual inferences rules. This stands in stark contrast to the tolerance paradox, which relies on counterfactual inference rules like rational monotonicity.

Going the other way, the sorites paradox requires inference rules that the tolerance paradox does not. In particular, the sorites paradox requires us to infer the material conditional on line four from the denial of sharp cutoffs on line three. But while that inference is classically valid, it is not intuitionistically valid—it is not valid if we go along with the intuitionists and deny the law of the excluded middle. The tolerance paradox, on the other hand, is intuitionistically valid. It goes through even if we deny the law of the excluded middle. So this is yet further reason to think that the paradoxes are genuinely distinct.

2.7. Lines

The tolerance paradox has a certain resemblance to a paradox from John J. L. Pollock (1976b). Suppose you draw a one inch line on a piece of paper. We then consider, how long would the line have been, had it been longer than an inch? Well, a world in which the line is less than two inches is more similar than one in which it is two inches. The familiar

Lewisian similarity semantics then tells us that had the line been longer than an inch, it would have been shorter than two inches. But then if space is continuous, the same goes for all the other lengths between two inches and one inch. Given **infinite agglomeration**, it then follows that had the line been longer than an inch, it would *not* have been longer than an inch.[13] So we have paradox.

There are important differences between our paradox and Pollock's line paradox. The first is that his paradox assumes that we are using something like Lewisian similarity semantics. Otherwise, there is no way to infer that (a) had the line been longer than an inch, it would have been shorter than two inches from (b) the fact that worlds in which the line is less than two inches are more similar to ours than worlds in which it is two inches. Our paradox, on the other hand, makes no assumptions about semantics. All we assumed was that counterfactuals were both tolerant and bounded and that certain rules of inference were valid. Now as a matter of fact, Lewisian similarity semantics validates the relevant rules, and so is subject to our paradox. But *any* semantics that validated those rules would be in the same position. There is no need to assume any particular connection between counterfactuals and similarity.

Second, the line paradox not only depends on using something like Lewisian similarity semantics. It depends on using what we might call offhand similarity as the similarity relation on worlds. Offhand, a line that is one and a half inches long is more similar to a line that is one inch long than a line that is two inches long. It must follow, then, that in the sense that matters for counterfactuals, a *world* in which a line is one and a half inches

13. $\wedge_i(A \mathrel{\Box\!\!\rightarrow} B_i) \supset (A \mathrel{\Box\!\!\rightarrow} \wedge_i B_i)$

long is more similar to our world than a world in which it is two inches long.

To his credit, Pollock recognizes that this is a substantial assumption and, in fact, suggests that we could resolve his paradox by denying it. Rather than using offhand similarity for our counterfactual semantics, we could use a specialized **coarse grained** similarity relation, one that counts worlds in which the line is between one inch and two inches, say, as equally similar. We then no longer get the result that had the line been longer than an inch, it would not have been longer than an inch.

There is much to be said for using a specialized similarity relation when doing similarity semantics. The important point for our purposes, though, is that while a coarse grained similarity relation might solve the line paradox, it gets us nowhere with the tolerance paradox.

Suppose that we use a similarity ordering that counts all the worlds in which Socrates is between six feet and seven feet as equally similar. This lets us make sense of the idea that had Socrates been at least six feet, he might have been at least a Planck length taller. So far so good. The problem now is the worlds at the top of the equivalence class—those worlds at which Socrates is exactly seven feet tall. All the worlds in which he is even a Planck length taller are in the next similarity class. So had Socrates been at least seven feet, he would not have been even a Planck length taller. We thus get a failure of tolerance. Coarse graining solves the problem only if we forget what the problem was in the first place.

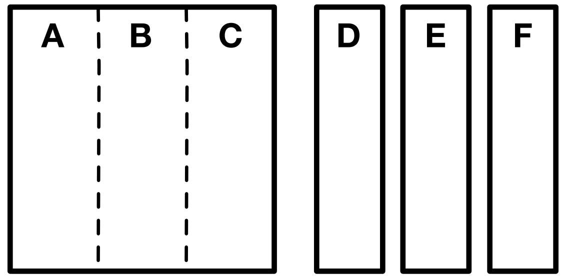
The third difference between the line paradox and ours is that his paradox relies on infinite agglomeration. One way of resolving his paradox, then, is to deny its validity. Lewis (1981) does exactly that: He accepts that for every length longer than an inch, had the line been longer, it would have been shorter than that. What he denies is that it follows that had the line been longer than an inch, it would not have been longer than an inch. The tolerance paradox, on the other hand, has nothing to do with agglomeration, nor does it use any infinitary inference rules. So the tolerance paradox and the line paradox are clearly quite different.

2.8. Woody

The tolerance paradox, as we have seen, is neither a version of the sorites paradox nor a version of the line paradox. What I want to suggest in this section is that there is in fact a much better analogy. The tolerance paradox, it seems to me, is a kind of counterfactual analogue of a certain well-known modal paradox from Chandler (1976) and Salmon (1979).

Consider a particular table named Woody. Ordinarily speaking, we think that things like tables could have been built out of slightly different matter. Had a few of the particles that compose Woody not existed, Woody could still have existed. His existence may depend on his matter, but it does not depend on *all* his matter. The composition of Woody is somewhat tolerant, we might say. That said, there are also limits to how different his matter could have been. Woody could have been built out of slightly different matter, sure. But he could not have been built out of *entirely* different matter.

Making all of this somewhat more precise, think of Woody as a tabletop rather than a table, one composed of three planks *ABC*. There are then three spare planks *D*, *E*, and *F* set to the side. Like this:



We then claim that Woody's composition is both modally tolerant in the following sense.

> **Modal Tolerance:** Necessarily, if Woody is made out of three planks, then he could have been made out of only two of those planks, together with some other third plank.
>
> **Modal Boundedness:** Necessarily, if Woody is made out of three planks, then he could not have been made out of three entirely different planks.

So for example, Woody is in fact made out of *ABC*. Modal tolerance then tells us that he could have been made out of *BCD*, since being made out of *BCD* means that Woody would have been made out of two actual planks and one spare plank. Modal boundedness says that he could not have been made out of *DEF*, since he would have been made out of entirely different planks.

Besides thinking that composition is tolerant and bounded, we also think that certain modal inferences preserve truth. For example, it would seem that:

S4 $\Diamond\Diamond A \supset \Diamond A$

Suppose that it could have been that Socrates was a possible fishmonger. In that case, it would seem to follow that Socrates is in fact a possible fishmonger. The possible possibilities are also possibilities, or so we generally think. Or equivalently, since possibility is the dual of necessity, we generally think that the necessities are themselves necessary. If Socrates is necessarily a human, then it is *necessary* that Socrates is necessarily a human. The propositions that are necessary is not itself a contingent matter.

Together, of course, these convictions lead to disaster. Because if constitution is modally tolerant and modally bounded and S4 is valid, we can prove a flat contradiction.

(1)	ABC	premise
(2)	$\neg\Diamond DEF$	modal boundedness
(3)	$\Box(ABC \supset \Diamond BCD)$	modal tolerance
(4)	$\Diamond BCD$	1,3, T
(5)	$\Box(BCD \supset \Diamond CDE)$	modal tolerance
(6)	$\Diamond\Diamond CDE$	4,5, K
(7)	$\Diamond CDE$	6, S4
(8)	$\Box(CDE \supset \Diamond DEF)$	modal tolerance
(9)	$\Diamond\Diamond DEF$	7,8, K
(10)	$\Diamond DEF$	9, S4
(11)	$\bot$	2, 10, PL

Besides S4, the paradoxical proof makes us e of classical logic and a pair of auxiliary modal axioms called T and K. But these are completely uncontroversial.[14] It looks, then, like we face a choice. We can deny modal tolerance, modal boundedness, or S4. That choice looks quite a bit like the choice we face in the counterfactual tolerance paradox.

14. T says that $\vdash \Box A \supset A$ and K says that $\vdash \Box(A \supset B) \supset \Box A \supset \Box B$. The first axiom corresponds to the idea that necessary truths are *truths*. The second, more or less, to the idea that a proposition is necessary iff it is true at all possible worlds.

There is of course an enormous literature on the Woody paradox and this is not the place to sift through all the possible responses. My own preferred response, though, is to deny S4. And how this is done in the modal case, it seems to me, suggests a way forward in the counterfactual case. This will be the topic of the next section.

2.9. Transitivity and Counterfactuals

Most readers will be familiar with Kripke models for modal logic. These are triples, much like the similarity models we saw in section 2.4.

$$\langle W, R, [\![\cdot]\!] \rangle$$

The main difference is that where the counterfactual accessibility relation is a three-place relation, the modal accessibility relation is just a two-place relation. But otherwise, the basic idea is surprisingly similar. We can then analyze the necessity and possibility operators in the usual way:

$w \models \Box A$ iff A is true at every $v \in W$ such that Rwv

$w \models \Diamond A$ iff A is true at some $v \in W$ such that Rwv

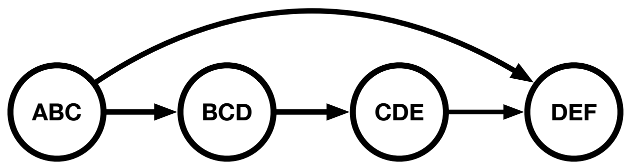
A sentence is necessary at a world when it is true at all accessible worlds. A sentence is possible at a world when it is true at some accessible world.

The observation now that there is a certain connection between the modal axiom S4 and the transitivity of the modal accessibility relation.

Transitivity: If Rwv and Rvu, then Rwu

In particular, the modal axiom is valid if and only if we require modal accessibility to be transitive.[15]

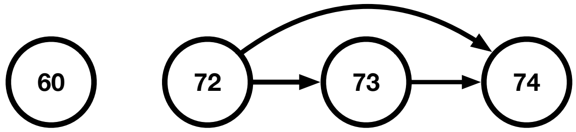
You can see the basic problem that transitivity raises. Suppose, as before, that we endorse both modal tolerance and modal boundedness. Given transitivity, modal space then ends up looking like this:



The actual world is one in which Woody is built out of ABC. Because he could have been built out of only two-thirds of his matter, he could have been made out of BCD. So BCD is accessible from ABC. But it is true at ABC that Woody *necessarily* could have been made out of two-thirds different matter. So CDE is accessible from BCD. Transitivity then tells us that it follows that CDE is accessible from ABC. But now, since CDE is accessible from ABC and it is necessarily possible at ABC that Woody is made out of only two-third of his matter, DEF is accessible from CDE. But then by transitivity again, DEF is accessible from ABC. So it is true at ABC that Woody could have been made out of entirely different matter. But that contradicts modal boundedness.

It turns out that transitivity cause more or less *the same problem* in the case of counterfactuals. For simplicity, think of counterfactual tolerance in terms of inches instead of Planck lengths. What we claim is that for any height greater than his actual height, had Socrates been at least that tall, he might have been at least one inch taller. Now suppose that we think about things in terms of similarity models. We then get the following:

15. Or more precisely, it is valid over the class of transitive frames.



Suppose that the actual world is on the left and that all of our similarity comparisons are fixed relative to the actual world. In that case, since it is true that had Socrates been at least 72 inches, he might have been 73 inches, it follows by our semantics that the 73 inch world is at least as similar as the 72 inch world. We represent that fact here using an arrow from the 72 inch world to the 73 inch world. But now by tolerance again, had Socrates been at least 73 inches tall, he might have been at least 74 inches tall. So the 74 inch world is at least as similar as the 73 inch world. But now by the *transitivity* of the similarity relation, the 74 inch world is at least as similar as the 72 inch world. But then by parity of reasoning, we can show that worlds in which Socrates is *arbitrarily* tall are at least as similar as the worlds in which he is 72 inches tall. But then by the Lewisian analysis of the might counterfactual, had Socrates been at least 72 inches, he might have been arbitrarily tall. So we get a contradiction of boundedness.

The problem, it would seem, is that *the similarity relation is transitive*. And so the solution would seem to be just as clear. We should keep the basic idea behind Lewisian similar similarity models, but give up transitivity.

2.10. Sufficient Similarity

What we are going to call **sufficient similarity models** are triples consisting of a set of worlds, a three-place accessibility relation on worlds, and a valuation function.

$\langle W, R, [\![\cdot]\!]\rangle$

These are models are in most respects just like the familiar Lewisian similarities models from section 2.4. The difference is that compared to Lewis, we are placing fewer requirements on the counterfactual accessibility relation. In particular, instead of requiring it to be total and *transitive*, we require it to be total and (weakly) acyclic.

Total Either $Rwvu$ or $Rvwu$

Acyclic If $\neg Rwvz$ and $\neg Rvuz$, then $\neg Rwuz$

As before, the counterfactual accessibility relation is just a set of triples, so could in theory represent any three-place relation between worlds you want, so long as it has the right formal properties. But one natural way to think about it is as representing facts about **sufficiently similarity**.

$Rwvu$ w is not sufficiently less similar than v to u

You can see that sufficient similarity is total and acyclic. That it is total means that there are never pairs of worlds w and v such that each is sufficiently less similar than the other. That it is acyclic means that if w is sufficiently less similar than v and v is sufficiently less similar than u, then w is sufficiently less similar than u. But now notice that while sufficient similarity is total and acyclic, it is not generally transitive. So we have a relation of the right form.

Once we have our sufficient similarity models, the counterfactual operators are defined the same way as before. We can just reuse the Lewisian clauses from section 2.4.

$u \models A \mathbin{\Box\!\!\rightarrow} B$ iff there is a world $v \in W$ such that $v \models A$ and every world $w \in W$ is such that $w \models A \supset B$ whenever $Rwvu$

$u \models A \mathbin{\Diamond\!\!\rightarrow} B$ iff for every world $v \in W$ such that $v \models A$, there is a world $w \in W$ is such that $w \models A \wedge \neg B$ and $Rwvu$

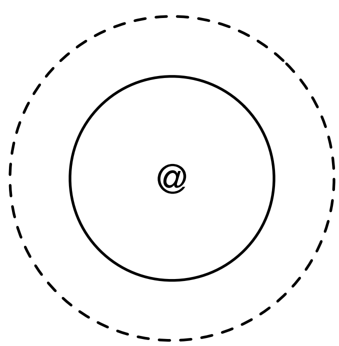
To simplify, suppose again that we have the limit assumption in place. What these clauses are then telling us is that:

$A \mathbin{\Box\!\!\rightarrow} B$ is true iff all of the sufficiently similar A worlds are B worlds

$A \mathbin{\Diamond\!\!\rightarrow} B$ is true iff some of the sufficiently similar A worlds are B worlds.

Here, the sufficiently similar A worlds are just the A worlds that are not sufficiently less similar than any other A world

Perhaps somewhat more intuitively, here is the contrast with Lewis. Lewis says that $A \mathbin{\Box\!\!\rightarrow} B$ is true iff all of the most similar A worlds are B worlds. Below, the most similar A worlds are represented by the solid sphere around the actual world.



To determine whether or not $A \mathbin{\Box\!\!\rightarrow} B$, then, we have to check to see if the solid sphere includes any A worlds that are also $\neg B$ worlds. The sufficient similarity approach, on the other hand, says that $A \mathbin{\Box\!\!\rightarrow} B$ is true iff all of the *sufficiently* similar A worlds are B worlds.

Here, the sufficiently similar A worlds are represented by the dotted sphere. As you can see, they include all of the most similar A worlds, but they also include A worlds that are just a bit less similar. So the region we need to check extends just a bit farther out into modal space.

Sufficient similarity models let us confirm that transitivity and rational monotonicity are in fact intimately related. That is, we can show that:

> Rational monotonicity is valid on precisely the class of transitive frames.

Thinking in terms of similarity, denying the validity of rational monotonicity is equivalent to denying the transitive of the counterfactual accessibility relation.[16] This confirms the suggested analogy with the Woody paradox. Rational monotonicity is to counterfactual logic as S4 is to modal logic. Both principles are valid when the accessibility relation is required to be transitive. And both principles fail once transitivity is denied.

---

16. The proof that rational monotonicity is valid on any transitive frames is basically just the proof from the footnote in section four. That rational monotonicity is valid on *only* transitive frames is also easy to show. Suppose that we have a non-transitive frame. This means that there are worlds $wvuz$ such that $Rwvz$ and $Rvuz$ but not $Rwuz$. Now consider a model with this frame in which $v(A) = \{w, v, u\}$ and $v(B) = \{v, u\}$ and $v(C) = \{v, u\}$. At world z, it is true that $A \mathrel{\Box\!\!\rightarrow} C$ and $A \mathrel{\Diamond\!\!\rightarrow} B$, but false that $A \wedge B \mathrel{\Box\!\!\rightarrow} C$. So we have our counterexample.

2.11. Selection

Perhaps the most prominent competitor to Lewisian semantics is the selection semantics from Robert Stalnaker.[17] What we are going to show in this section is that the state of play is much the same for selection semantics as it is for Lewisian similarity semantics. If we use precise similarity, and so have a transitive similarity relation, then we have a paradox. But if we use a non-transitive similarity relation instead, the paradox is resolved.

Stalnaker thinks about counterfactuals in terms of **selection models**. These are triples consisting of a set of worlds, a selection set, and a valuation function.

$$\langle W, S, v \rangle$$

The selection set is a set of selection functions, each of which maps propositions and worlds to worlds.[18] Think of these functions as each representing an "opinion" about which worlds are closest.

$$f(A, u) = w \quad \text{the closest } A \text{ world to } u \text{ is } w$$

Moreover, not only does each selection function have an opinion about which worlds are *closest*. They also have an opinion about which worlds are *closer*. That is, each selection

17. See for example Stalnaker (1968), Stalnaker and Thomason (1970), and Stalnaker (1981).

18. More precisely, a selection function is an $f : \mathcal{P}(W) \times W \to W$. The "propositions" here are thus sets of worlds. A propositions is "true at a world" when it has that world as a member. There are then two further constraints. First, $f(A, w) \in A$. The closest A world is always an A world. Second, if $f(A, w) \in B$ and $f(B, w) \in A$, then $f(A, w) = f(B, w)$. If the closest A world is a B world and the closest B world is an A world, then those worlds are the same.

function determines a unique closeness ordering on worlds.[19] That ordering is strict in the sense that it has no ties. As far as selection functions are concerned, no two worlds are ever equally close. Finally, that selection models have a *set* of selection functions, rather than just a single selection function, is meant to correspond to the idea that closeness is generally somewhat vague or otherwise underdetermined.

Once we have selection models, the counterfactual operators are defined by supervaluating over the selection functions in the selection set. This means that we need to first define truth relative to a selection function and a world.

$f, w \models A$	iff $w \in v(A)$, when A is atomic
$f, w \models \neg\phi$	iff $f, w \not\models \phi$
$f, w \models \phi \vee \psi$	iff $f, w \models \phi$ or $f, u \models \psi$
$f, w \models \phi \;\Box\!\!\rightarrow \psi$	iff $f(\phi, w) \models \psi$
$f, w \models \phi \;\Diamond\!\!\rightarrow \psi$	iff there is an $f^* \in S$ such that $f^*(\phi, w) \models \psi$

We then say that a sentence is true at a world full stop when it is true at that world relative to all selection functions.[20]

So far so good. We can now establish a certain connection between selection models and similarity relations. In particular, say that a selection function is **compatible** with a similarity relation when:

$$f(A, u) = w \text{ only if } Rwvu$$

19. A world w is closer than v to u according to the selection function f iff there is an $A \subset W$ such that $w, v \in A$ and $f(A, u) = w$

20. So $w \models \phi$ iff $f, w \models \phi$ for all $f \in S$.

A selection set S and a similarity relation R then **correspond** when S is the set of all and only the selection functions that are compatible with R. What we can show is that each selection set corresponds to a unique similarity relation, and vice versa.

This means that we can think of selection models as offering a *similarity* analysis of counterfactuals. It's just that facts about similarity are coded up using a set of selection functions rather than an explicit similarity relation. The natural question then is: What sort of views about similarity are we committing ourselves to when we use selection models?

The perhaps surprising answer is that we are committing ourselves to the idea that *similarity is non-transitive*. To see why, consider a simple model. It has four worlds and two selection functions. The selection functions then rank the closeness of worlds to world z as follows:

f_1	f_2
w	v
u	w
v	u
z	z

Worlds that are strictly closer to z are lower on the table. Worlds that are strictly farther from z are higher on the table. Now given a selection set, the corresponding similarity relation is the one such that:

$Rwvu$ iff there is some $f \in S$ according to which w is closer to v than u

But now looking at the above table, you can see that the corresponding similarity relation is not transitive. There is a selection function according to which w is closer to v and there is a selection function according to which v is closer than u. So $Rwvz$ and $Rvuz$. But there is no selection function according to which w is closer than u. So not $Rwuz$.

Now of course, we can force selection sets to correspond to transitive similarity relations by imposing further requirements. We can require them be **regular** for example.[21] If we do that, transitivity will be regained. Every *regular* selection set corresponds to a transitive similarity relation. This despite the fact that selection sets in general need not correspond to transitive similarity relations.

As far as I can tell, the standard view is that if we are going to use selection models, then we should assume regularity. We should assume that the selection functions together encode a transitive similarity relation. The question, though, is whether this standard view can be maintained in the face of the tolerance paradox. And the answer, I think, is that it cannot.

Now there is a certain hope that once we make the move to selection models, transitivity can be maintained. The reason is that rational monotonicity is invalid *even if* we require selection sets to be regular and so, in effect, require the corresponding similarity relation to be transitive.[22] This means that our original paradoxical argument is invalid. So it looks

---

21. A selection set S is regular iff whenever there is an $f_1 \in S$ according to which w is closer than v and there is an $f_1 \in S$ according to which v is closer than u and $w \neq u$, then there is an $f_1 \in S$ according to which w is closer than u. We are calling this principle regularity because it plays a functionally similar role to a principle that van Fraassen (1974) calls regularity.

22. Consider any model such that $W = \{w, v, u\}$ and $\{f_1, f_2\} \subset S$ and $A = \{w, v, \}$ and $B = \{v\}$ and $C = \{w\}$. Moreover, suppose that $f_1(A, u) = w$ and $f_2(A, u) = v$ and $f_1(A \wedge B, u) = f_2(A \wedge B, u) = v$. In that case, $(A \,\Box\!\!\rightarrow C) \wedge (A \,\Diamond\!\!\rightarrow B) \supset (A \wedge B \,\Box\!\!\rightarrow C)$ is false at u relative to f_1.

like selection models give us a way to maintain the idea that similarity is transitive while also resolving the paradox.

Unfortunately, this hope is misplaced. Regular selection models may invalidate rational monotonicity. But they also validate a different principle called **restricted might transitivity.**[23]

RMT $(A \diamondrightarrow B) \wedge (A \wedge B \diamondrightarrow C) \supset A \diamondrightarrow C$

The problem then is that once we have have this principle, we can run a revised version of the paradox without rational monotonicity. First, we reformulate boundedness.

> **Boundedness** ($\diamondrightarrow$): There are j and k such that it is false that had Socrates been at least h, he might have been at least k.

Instead of affirming that Socrates would not have been arbitrarily tall, we simply *deny* that he might have been arbitrarily tall. We then reason to contradiction much like before.

23. Suppose that $A \diamondrightarrow B$ and $A \wedge B \diamondrightarrow C$ are both true at world z and fix all similarity and closeness comparisons to that world. So there are closeness relations in S such that the closest A world is a B world and the closest $A \wedge B$ world is a C world. By compatibility, it follows that there is an $A \wedge B$ world w that is at least as similar as any other A world and an $A \wedge B \wedge C$ world v that is at least as similar as any other $A \wedge B$ world. But from this it follows that w and v are equally similar. So by the transitivity of equisimilarity, it follows that v is as similar as any other A world. But this means that there is a closeness relation such that v the most similar A world. Meaning that the there is a closeness relation on which the closest A world is a C world. So $A \diamondrightarrow C$ is true at z.

(1) $\neg(j \diamondto k)$ boundedness ($\diamondto$)

(2) $j \diamondto j+1$ tolerance

(3) $n \diamondto n+1$ hypothesis

(4) $n+1 \diamondto n+2$ tolerance ($\diamondto$)

(5) $n \wedge (n+1) \diamondto n+2$ 4, substitution

(6) $n \diamondto n+2$ 3,5, RMT

(7) $j \diamondto k$ 2,6, induction

(8) $\bot$ 1, 7, PL

Regular selection models, then, may invalidate rational monotonicity, but they do not resolve the paradox.

The solution to the paradox is to deny regularity. This invalidates both rational monotonicity and restricted might transitivity, so invalidates both version of the paradox. But denying regularity means denying transitivity. It means using selection sets that correspond to sufficient similarity relations instead of precise similarity relations. And so the solution for followers of Stalnaker is the same as the solution for followers of Lewis. In both cases, the transitivity of similarity is what has to go.

2.12. Disconfirmation

Building sensible models on which rational monotonicity can fail is the first step towards resolving the paradox. To have a full resolution, though, we need to explain not just how the inference can fail, but why we mistakenly found it so compelling in the first place.

Something we would like from a theory of counterfactuals is a theory of disconfirmation.

We like an explanation of how we can—and often do—convince each other to give up certain counterfactual opinions. For example, suppose I claim that:

> Had Naomi played in the tennis tournament, she would have won.

You, on the other hand, disagree. You say that:

> Had Naomi played in the tennis tournament, she might have lost.

Clearly, there is some sense in which the two of us disagree. The sense in which we disagree, though, will depend on your theoretical commitments. For Lewis, our disagreement is a straightforward disagreement about matters of fact. I believe one thing and you believe something else that is logically inconsistent with what I believe. For Stalnaker, the conflict will be a bit more subtle. Perhaps I take myself to *know* that had Naomi played in the tennis tournament, she would have won. But you think that this is going too far. You think that for all we know—are perhaps for all anyone could ever know—Naomi would have lost.

However we understand the exact nature of our disagreement, what should be clear is that there are certain strategies you might use in order to change my mind. You might, for example, say something like the following:

> Look, I can see where you're coming from. But had Naomi played in the tournament, she might have faced Serena. And had she played in the tournament and played Serena, she would have lost. And so what you say is simply not correct. Had Naomi played in the tournament, she might have lost.

This sort of reasoning would seem to be clearly valid. If you're right that (a) Naomi might have faced Serena and that (b) in that case, she would have lost, then (c) it follows that

Naomi might have lost.

The question is, what backs this sort of reasoning? One possibility is that the reasoning is backed by a rule like strengthening.

ST $(A \mathrel{\Box\!\!\rightarrow} C) \supset (A \wedge B \mathrel{\Box\!\!\rightarrow} C)$

That is, whenever you have a true would counterfactual, you can get another true would counterfactual just by conjoining a proposition to the antecedent. In that case, my original claim entails that had Naomi played in the tournament *and played Serena*, she would have won. But if you're right that Naomi would have lost if she had played Serena, then this is clearly false. So my original claim must be false too.

The problem is that strengthening is invalid, as we learned from Lewis. If Kangaroos had no tails, they would have toppled over. But if they had advanced prosthetics in place of tails, they would not have toppled over. Had Naomi gone to the party, she would have had a good time. But had Naomi and Sam both gone to the party, Naomi would not have had a good time. But had Naomi and Sam and Serena all gone to the party, Naomi would have had a good time. And so on and so forth.

Lewis's solution to failure of strengthening is to put rational monotonicity in its place. Rational monotonicity is a kind of restricted strengthening principle. It says that we can strengthen from $A \mathrel{\Box\!\!\rightarrow} C$ to $A \wedge B \mathrel{\Box\!\!\rightarrow} C$ under certain special circumstances. Which circumstances are those? The ones in which $A \mathrel{\Diamond\!\!\rightarrow} B$. This lets us explain why, for example, it can be true that kangaroos would have toppled over if they had no tails, but would not have toppled over if they had advanced prosthetics. It can be true because it is *false* that if kangaroos had no tails, they might have had advanced prosthetics.

If rational monotonicity were valid, it would explain the validity of your reasoning. After all, I claim that had Naomi played in the tournament, she would have won. You point out that had she played, she might have faced Serena. So it follows by rational monotonicity and my original claim that had Naomi played in the tournament and faced Serena, she would have won. But we agree that this is false. So on pain of inconsistency, I have to retract my original assertion.

The problem, of course, is that rational monotonicity leads directly to the paradox of counterfactual tolerance. So what we need is an explanation of how you can go about changing my mind that does not depend on rational monotonicity.

What I think is that our counterfactual practice is backed by a rule you might call disconfirmation.

DC $(A \diamond\!\!\rightarrow B) \wedge (A \wedge B \,\Box\!\!\rightarrow C) \supset (A \diamond\!\!\rightarrow C)$

This rule directly validates your earlier reasoning. Had Naomi entered the tournament, she might have faced Serena. Had she entered the tournament *and* faced Serena, she would have lost. So it follows that had Naomi entered the tournament, she might have lost–and this precisely because she might have faced Serena.

The advantage of disconfirmation is that it lets us steer clear of the paradox. We know this because disconfirmation is valid over the class of all near similarity models.[24]

24. Suppose $A \wedge B \,\Box\!\!\rightarrow C$ at z and fix all near similarity comparisons to z. This means that there is an $A \wedge B$ world w such that all the $A \wedge B$ worlds that are at least nearly as similar are C worlds. Now suppose for reductio that $A \,\Box\!\!\rightarrow \neg C$. There is thus an A world v such that all of the A worlds that are at least nearly as similar are $\neg C$ worlds. So in particular, w is not at least nearly as similar as v. Next, suppose that $A \diamond\!\!\rightarrow B$

It is also valid over the class of all selection models.[25] But since rational monotonicity is invalid on both classes, it follows that we can accept disconfirmation while denying rational monotonicity.

Moreover, you can see where falsification might easily be *confused* with rational monotonicity, and so explain why we found rational monotonicity so compelling in the first place. Given duality and the counterfactual law of the excluded middle, disconfirmation and rational monotonicity are logically equivalent. For followers of Lewis, this means that whatever explains our (mistaken) attraction to the counterfactual law of the excluded middle can also explain our (mistaken) attraction to rational monotonicity. For followers of Stalnaker, something similar is true. Whatever explains our (mistaken) attraction to duality can also explain our (mistaken) attraction to rational monotonicity. In both cases, we can just reduce the one mistake to another mistake we are already committed to explaining.

---

at z. From this it follows that there is an $A \wedge B$ world u that is at least nearly as similar as v. But then since all of the $A \wedge B$ worlds that are at least nearly as similar as v are $\neg C$ worlds, it follows that u is a $\neg C$ world. This means that u is not at least nearly as similar as w. But we already said that w is not at least nearly as similar as v. So by acyclicality, it follows that u is not at least nearly as similar as v. But this is contrary to assumption. So $\neg(A \mathrel{\Box\!\!\rightarrow} \neg C)$ at z and therefore $A \mathrel{\Diamond\!\!\rightarrow} C$ at z.

25. Suppose that there is some $f \in S$ according to which the closest A worlds is a B world. Suppose, moreover, that the closes $A \wedge B$ world is a C world according to every $g \in S$. It then follows that the closest $A \wedge B$ world is a C world according to f. But given the constraints governing selection functions, this means that the closest A world according to f is identical to the closest $A \wedge B$ world. So there is some $f \in S$ such that the closest A world is a C world.

2.13. Pollock

John Pollock suggests that we should analyze counterfactuals using what we are going to call **Pollock models**.[26] His models invalidate rational monotonicity, so would seem to offer a way to resolve the tolerance paradox. In this section, I want to say why I think that sufficient similarity models are preferable.

Like sufficient similarity models, Pollock models are triples consisting of a set of worlds, a counterfactual accessibility relation, and a valuation function.

$$\langle W, R, v \rangle$$

There are then two key differences. The first is that Pollock does not require his accessibility relation to be total and acyclic. Instead, he requires it to be reflexive and transitive.

Reflexive	$Rwwz$
Transitive	if $Rwvz$ and $Rvuz$, then $Rwuz$

The second difference is that Pollock defines the counterfactuals somewhat differently. When we constructed sufficient similarity models, we followed Lewis and used the first of the below definitions. Pollock uses the second. The difference when it comes to might counterfactuals is similar.

$u \models A \mathrel{\Box\!\!\rightarrow} B$ iff there is a world $v \in W$ such that $v \models A$ and every world $w \in W$ is such that $w \models A \supset B$ whenever $Rwvu$

26. See for example J. L. Pollock (1975, 1976b, 1976a, 1981).

$u \models A \mathrel{\Box\!\!\rightarrow} B$ iff there is a world $v \in W$ such that $v \models A$ and every world $w \in W$ is such that if $w \models A \wedge \neg B$, then not both $Rwvu$ and not $Rvwu$

As it turns out, these two difference cancel out and the result is a kind of formal equivalence. Pollock models and sufficient similarity models validate all the same rules and axioms. So from a purely formal perspective, I have no beef with Pollock.

The problem is making sense of his accessibility relation. Pollock thinks of his accessibility relation as the **containment of change** relation.

Rwvu the changes needed to get from u to w are a subset of the changes needed to get from u to v

What this means is that when *Rwvu* and not *Rvwu*, the changes needed to get from u to w are a *strict superset* of the changes needed to get from u to v. To get from world u to world w, you have to do everything you need to do to get from u to v, and then some.

For my own part, I am not sure that I understand the containment of change relation. And insofar as I do, I strongly doubt that it has the right extension. For example, the containment of change relation would seem to stand in the way of solving the tolerance paradox. Here is why: Suppose that Socrates is in fact five feet tall. Now consider any pair of worlds w_n and w_{n+1} that are otherwise similar, but at which Socrates is n and $n+1$ Planck lengths taller, respectively. Insofar as I understand the containment of change relation, the changes needed to get from the actual world to w_{n+1} are a strict superset of the changes needed to get from the actual world to w_n. To make Socrates $n+1$ Planck lengths taller, you have to do everything you have to do to make him n Planck lengths taller, and then some, with the same going for every other such pair of worlds. But in that case, tolerance fails. Had Socrates been at least six feet tall, he would not have been even a Planck length taller.[27]

What this shows, it seems to me, is that we should scrap Pollock's containment of change relation and replace it with something else. But what? One idea would be to go back to using precise similarity. We could say that:

$Rwvu$ w is at least as similar as v to u

Failures of totality would then correspond to failures of comensurability. Sometimes, there are pairs of worlds w and v such that w is neither strictly more similar, nor strictly less similar, nor equally similar.

The problem is that failures of rational monotonicity will then require failures of commensurability. So for example, to block the tolerance paradox, we will need pairs of worlds w_n and w_{n+1} whose similarity to the actual world is incommensurable. But why would they be *incommensurable*? You would think that w_n would be strictly more similar than w_{n+1}. Or perhaps they should count as equally similar because a single Planck length is too small to make a difference. Either way, though, their similarity is commensurable. What we need is some reason to think that their similarity is *incommensurable*. And I have no idea what that could be.

27. A similar point applies to Kratzer (1981). She suggests a variation of Pollock's semantics on which the containment of change relation is fixed using what she calls an ordering source. But while her semantics validates the same rules and axioms as sufficient similarity models, the problem is finding a natural order source that gives us both tolerance and boundedness. For my own part, at least, I have been unable to find one.

Maybe if we fumbled around long enough, we could save Pollock models. We could find a relation with both the right formal properties and the right extension. A simpler solution, though, is to switch to sufficient similarity models. Sufficient similarity models validate all the same rules and axioms. But in the case of sufficient similarity models, we already have a relation that would seem to have both the right formal properties and the right extension. So why bother?

2.14. Independence

Pollock's own reasons for denying rational monotonicity are not especially compelling. The cases in his J. L. Pollock (1976b) and J. Pollock (1981) are somewhat baroque, so we can instead consider a simpler version from Boylan and Schultheis (2018).

Suppose we have three characters Alice, Bernie, and Carole who flip three fair coins. Those flips are causally independent. As it happens, Alice and Carole both flip heads. Bernie flips tails. We then consider three counterfactuals:

(1) Had Alice and Bernie flipped the same, Carole would have flipped heads.

(2) Had Alice and Bernie flipped the same, Alice and Bernie and Carole might have all flipped the same.

(3) Had Alice and Bernie and Carole flipped the same, they would have all flipped heads.

The first two are true, but the third is false, or so you might think. Why is that? Well, suppose you accept a certain plausible principle connecting counterfactuals and causation.

Independence: If $\neg A$ and B are both true at the actual world, then $A \mathrel{\Box\!\!\rightarrow} B$ iff those facts are causally independent.

This principle then entails that the first two counterfactuals are true and that the third is false. So rational monotonicity is invalid.

Here is how the reasoning goes: The first counterfactual is true because changing how Alice and Bernie flipped would not have changed how Carole flipped, and she in fact flipped heads. The second is true because had Alice and Bert flipped the same, that might have been because Bert flipped heads. And by independence again, Carole would still have flipped heads. So they all might have flipped heads and so might have all flipped the same. But the third counterfactual is false because all three flipping the same is not causally independent from all three flipping heads. After all, all three flipping heads *just is* a way of all three flipping the same.

The problem for Pollock is that independence-style reasoning not only predicts the invalidity of rational monotonicity. It also predicts the invalidity of other principles like **restricted strengthening** that are validated by his semantics.[28] So if we are going to deny rational monotonicity on the basis of independence, Pollock's semantics is not the way to

28. Restricted strengthening says that $(A \mathrel{\Box\!\!\rightarrow} B) \wedge (A \mathrel{\Box\!\!\rightarrow} C) \supset (A \wedge B \mathrel{\Box\!\!\rightarrow} C)$. The observation that independence entails the failure of restricted strengthening is from Walters (2009).

do it. We are going to need a much more radical departure from Lewisian semantics.[29,30]

The advantage of the tolerance paradox is that we can motivate the denial of rational monotonicity without appealing to anything like causal independence. And so we can just sidestep the whole problem.

2.15. Dynamic Semantics

One especially simple view is that would counterfactuals are strict conditionals. $A \mathrel{\Box\!\!\rightarrow} B$ is equivalent to $\Box(A \supset B)$. Might counterfactual are then understood as expressing certain kinds of possibilities. In particular, $A \mathrel{\Diamond\!\!\rightarrow} B$ is equivalent to $\Diamond(A \wedge B)$. Like Lewisian semantics, the strict conditional view validates all the rules used in the tolerance paradox.[31] This means that defenders of the strict conditional view will have to either deny that counterfactuals are tolerant or deny that they are bounded.

That may be. But nowadays, defenders of the strict conditional often accept certain dynamic mechanisms for shifting contexts. To say that $\Box(A \supset B)$ is to say that $A \supset B$ is true at all the relevant worlds. Since the relevant worlds can shift from context to context, the truth of counterfactuals can shift from context to context. This gives defenders of the strict

29. I owe this point to Simon Goldstein.

30. You could think that the relevant judgements are correct, but deny that this on the basis independence. Myself, I have a hard time hearing the first two sentences as true and the third as false. And when I can get those readings, I strongly suspect that I'm simply smuggling in independence-style reasoning.

31. More precisely, it validates substitution, duality, and rational monotonicity so long as $\Box$ is a normal modal operator and $\Box\phi$ is equivalent to $\neg\Diamond\neg\phi$.

conditional analysis additional powerful tools for explaining away counterexamples. You might think, then, that these sorts of tolls might help with the tolerance paradox.

To see how context shifts can help, consider apparent counterexamples to strengthening, an especially potent cousin of rational monotonicity that we met in section 2.12. Where rational monotonicity says that we can strengthen from $A \mathrel{\Box\!\!\rightarrow} C$ to $A \wedge B \mathrel{\Box\!\!\rightarrow} C$ under certain special circumstances, strengthening says that we can strengthen under any circumstances whatsoever. Strengthening is validated by the strict conditional analysis, but would seem to have clear counterexamples. To use the case from Lewis again, if kangaroos had no tails, they would have toppled over. But if kangaroos had advanced prosthetics in place of tails, they would not have toppled over.

One strategy for explaining away such counterexamples is suggested by Fintel (2001). Here is the basic idea: Suppose we start in a context in which the only relevant possible worlds in which kangaroos have no tails are worlds in which they topple over. Furthermore, there are no relevant worlds in which they have advanced prosthetics. You then assert that if kangaroos had no tails, they would have toppled over. This is true, given the strict conditional analysis. Now you go on to assert that if kangaroos had advanced prosthetics in place of tails, they would not have fallen over. This assertion, von Fintel thinks, presupposes that the context includes worlds in which kangaroos have advanced prosthetics. Since the original context has no such world, the context shifts to accommodate. In particular, it shifts by adding all of the most similar worlds in which kangaroos have advanced prosthetics to the set of contextually relevant worlds. Since those are worlds in which kangaroos do not topple over, your second assertion is true.

So far so good. The question now is whether a similar mechanism can explain away the tolerance paradox. To fix on an example, suppose that we start out in a context that includes worlds in which Socrates is up to seven feet tall. I then assert that had Socrates been at least six feet tall, he would have been no taller than seven feet. This true, given the context, so we get boundedness. The question is whether we can explain tolerance. Suppose I claim that had Socrates been at least seven feet tall, he might have been a Planck length taller. This is obviously true, but is false on the strict conditional analysis, since the context does not include any worlds in which Socrates is taller than seven feet.

One idea, suggested by Gillies (2007), is that asserting a might counterfactual $A \diamond\!\!\rightarrow B$ presupposes that the context includes worlds at which $A \wedge B$. If the context has no such worlds, it accommodates the presupposition by adding the most similar $A \wedge B$ worlds. This is how that might help with the previous example: I assert that had Socrates been at least seven feet, he might have been at least a Planck length taller. There are no worlds where Socrates is a Planck length taller than seven feet in the original context, so the context shifts to accommodate by adding the most similar such worlds. My assertion is then true in the new context. So it would seem, we have a dynamic strategy for explaining away the tolerance paradox. Counterfactuals seem tolerant because contexts always accommodate might counterfactuals.

A serious problem with the proposed mechanism is that it is overly accommodating. If I say that had Socrates been at least seven feet, he might have been at least eight feet, the context will shift to accommodate. Sure. That's fine. But for the same reason, if I say that had Socrates been at least seven feet, he might have been a million feet, the context will *also* shift to accommodate. But this is the wrong result. Had Socrates been at least seven feet, he would not have been a million feet. Or take another example. Suppose you make the following speech:

> Had Socrates been at least a million feet, he might have been a million feet. But had he been at least six feet, he would have been less than ten feet.

The first assertion presupposes that there is a contextually relevant world in which Socrates is a million feet tall. So if the context has no such world, it accommodates by adding one. But in that case, your second assertion is false. But again, this is the wrong result. Saying that had Socrates been a million feet, he would have been a million feet should have no bearing on whether you can go on to say that had Socrates been at least six feet, he would have been less than ten feet.

Perhaps we should not read too much into such difficulties. These are only objections to a particular mechanism for shifting contexts, and there may well be more sophisticated mechanisms that can do the job. Still, to actually solve the tolerance paradox, we would need to specify those mechanisms. And it is surprisingly hard to see what they might be.[32]

2.16. Conclusion

The first half of this paper presented the paradox of counterfactual tolerance and the second presented my own preferred solution. I think that what the paradox shows is that we need to analyze counterfactuals using a non-transitive similarity relation. We need to use sufficient similarity in place of precise similarity.

All of that said, I am more committed to the *paradox* than I am to any particular

32. Similar skepticism about dynamic semantics is raised in Boylan and Schultheis (2018).

solution. What I think is that any plausible theory of counterfactuals will have to resolve the paradox one way or another and that—however the paradox is resolved—we will have to say something interesting. The paradox thus serves as an important constraint on what our final theory of counterfactuals can look like.

3 Multi-Dimensional Quantified Modal Logic

The most familiar modal logics are one-dimensional in the sense that they are characterized by models in which sentences are evaluated at individual worlds. But while one-dimensional modal logics may be most familiar, they also have certain limitations. And those limitations have convinced many of us that important modal notions—like metaphysical necessity and physical necessity—cannot be fully characterized in a one-dimensional framework. What is needed, then, is a multi-dimensional framework, one in which sentences are evaluated at *sequences* of worlds rather than just worlds.

In chapter one, we saw what I think of as one of the most interesting applications of multi-dimensional modal logic. There, we used a multi-dimensional language called relational modalese to build a modal theory of distance ratios. This was meant to serve as a kind of proof of concept: If the basic physical facts include facts about how things could have differed—and not just how things could have been—then we have a powerful tool for doing science with only particles.

The goal of this chapter is to convince you that multi-dimensional modal logic is in good order. In particular, we are going to revisit the motivations for multi-dimensional modal logic in section one. We saw many of these in chapter one, but it is worth seeing them again from a more technical perspective. We are then going to specify a hierarchy of propositional modal languages (PMLs) in section two and a corresponding hierarchy of quantified modal languages (QMLs) in section three. Along the way, we are going to generalize the familiar notion of a Kripke model to arbitrarily many dimensions. This is meant to correct what I see as a certain problem in the literature. There, the most commonly used multi-dimensional models validate all the axioms of S5. Even when more general models are used, they still validate axioms that are less than ideal in certain contexts. For example:

$$\phi \supset \Diamond\phi$$

But now suppose that we are interested in building a multi-dimensional *deontological* logic, and so want to think of the diamond operator as expressing permission. In that case, this tells us that whatever is actual is also permissible, which is clearly absurd. And so if multi-dimensional modal logic is to have its widest range of application, we are going to need a more general framework.

After describing a hierarchy of languages and models, we are going to turn our attention to proof theory. For simplicity and concreteness, we are going to focus on the two-dimensional case, but the basic approach extends to higher dimensions as well. Section four describes several systems, the most important of which are $\mathbf{K_2}$ and $\mathbf{Q_2}$. The first is the logic of all two-dimensional Kripke models. The second is the logic of all *quantified* two-dimensional Kripke models. That this is so will be shown in sections five and six, where we provide completeness results for both systems. To my knowledge, there are no published completeness results for quantified multi-dimensional modal logic. And so this second result is meant to fill what I see as an important gap in the literature.

3.1. Motivation

The language of quantified predicate logic (QPL) was built to regiment the practice of mathematics. But the practice of mathematics, being what it is, is a purely categorical enterprise. It is about how certain mathematical structures are, not how they *could* have been, *should* have been, or *were*. Ordinarily, though, we think that there are facts involving such notions. We think that there *could* have been more planets, that there *should* have been less suffering, and that there *were* once dinosaurs who roamed the earth. This then raises a natural question: Given that there are such facts, what sort of formal language should we use to express them?

There are two basic approaches. The **regimented** approach says that despite initial appearances, the language of quantified predicate logic is the right one after all.[1] All we need is quantification over the right sort of indexes. In the modal case—which will be our main focus here—those indexes are possible worlds.[2] To fix on an example, proponents of

---

1. My use of the regimented/autonomous terminology is borrowed from Burgess (2009). The distinction is similar to—but also distinct from—the modalist/possiblist distinction from chapter one. In particular, the autonomous/regimented distinction is purely linguistic. Neither view makes any particular claims about the nature of reality. The modalist/possiblist dispute, on the other hand, is meant to be metaphysical. The modalist thinks that modality is ultimately about how things could have been. The possiblist thinks that modality is ultimately about what there is. Generally speaking, modalists are aligned with the autonomous approach and possiblists with the regimented approach. But you could also imagine various non-standard ways in which the views might be combined.

2. For simplicity, we are going to generally think of the regimented approach as an *indexing* approach,

the regimented approach might use a two-sorted quantified predicate language $\mathcal{E}$ with one stock of variables w, v, u... for possible worlds and another x, y, z... for the individuals who inhabit those worlds. Names and argument places are then similarly sorted. The logical predicates of the language are:

Aw	w is the actual world
Ixw	x is in w
$x = y$	x is identical to y

The actuality predicate takes terms for worlds. The in predicate takes terms for individuals in the first position and terms for worlds in the second. The identity predicates takes terms of either sort in both positions. Modal facts are then expressed by indexing non-logical predicates to worlds. So for example, we could say that there is a possible world at which Socrates is a farmer by writing $\exists w Fsw$.

The second approach is the **autonomous** approach. The autonomous approach says that intensional notions are best expressed by extending the language of QPL with certain sentential operators. In the modal case, the standard autonomous language is the language of one-dimensional quantified modal logic, which we are going to call Q_1. That language is built by adding a necessity operator $\Box$ to the language of QPL and then defining a possibility operator $\Diamond$ in terms of it. This lets us express the idea that Socrates could have been a farmer by writing $\Diamond Fs$.

---

meaning that it quantifies over possible worlds and adds argument places to predicates. There are also other strategies, though, that you might pursue. You might, for example, follow Lewis 1968 and use quantification over counterparts instead of quantification over worlds.

In many cases, we can **directly translate** between the regimented language $\mathcal{E}$ and the autonomous language Q_1, For example:

$$\exists w \exists x(Ixw \wedge Fsw) \mapsto \Diamond \exists x Fx$$

If I say that there is a possible world at which someone is a farmer, you can translate that by saying that it could have been that someone was a farmer. This is a direct translation in the sense that it directly trades quantification over worlds for operators. You can even see how the translation might be the result of a certain purely *mechanical* (and in fact recursive) procedure. That is, to get from the first sentence to the second, we just (a) exchange the existential quantification over worlds for a diamond, (b) exchange the restricted quantification over individuals for unrestricted quantification, and (c) delete the the extra argument place from the *is a farmer* predicate. This sort of procedure could be made fully precise. But you get the basic idea.[3]

The problem for the autonomous approach is that there are natural modal claims that can be expressed in $\mathcal{E}$ that have no direct translation into Q_1. One kind of familiar example is from Hazen (1976). Ordinarily, we think that:

There could have been something that does not actually exist.

This fact can be easily expressed in the regimented language. We can just say that:

3. These sorts of direct translations are sometimes called *standard translations* and are the subject matter of correspondence theory. For more on this, see the second chapter of Blackburn, Rijke, and Venema (2010). For some examples of multi-dimensional standard translations, see Marx and Venema (1997).

$$\exists v(Av \wedge \exists w \exists x(Ixw \wedge \forall y(Ixv \supset x \neq y)))$$

The actual world is v and there is a possible world w such that some x in w is not identical to anything in v. The problem is that there is no way to directly translate this sentence into $\mathcal{Q}_1$, and so no way to directly express the idea that there could have been something that does not actually exist.

A second kind of case involves relations across worlds. Suppose, for example, that Aristotle is taller than Socrates. This fact is naturally expressed in quantified predicate logic by writing Tas. But now suppose that we want to say that:

> Aristotle could have been taller than Socrates actually is.

This is naturally expressed in the regimented language by using a double-indexed *taller than* predicate:

$$\exists v(Av \wedge \exists w(Tawsv))$$

The actual world is v and that there is a possible world w such that Aristotle at w is taller than Socrates at v. The problem for the autonomous approach is that again, there is no way to directly translate this perfectly sensible claim into $\mathcal{Q}_1$.

As it turns out, these two sorts of problems are picking up on the same structural feature. Namely: $\mathcal{Q}_1$ can only translate sentences of $\mathcal{E}$ that have a single world variable.[4]

---

4. While this is *mostly* true, there is also an exception involving the accessibility relation. Suppose, for example, that we add an accessibility predicate to $\mathcal{E}$. If our modal operators are also appropriately restricted by accessibility, we can then translate $\exists v(Av \wedge \exists w(Rwv \wedge Fsw))$ as $\Diamond Fs$. But this sentence requires two world variables so, strictly speaking, it is not true that $\mathcal{Q}_1$ cannot translate such sentences. What is true is that it

But both of the above claims require at least *two* world variables. And so there is no way of directly translating them into $\mathcal{Q}_1$. Or putting it somewhat more intuitively: $\mathcal{Q}_1$ is a language for describing the intrinsic features of worlds. But ordinarily, we think that there are certain kinds of *comparisons* and *relations* between worlds. These are easy to describe in $\mathcal{E}$. But those descriptions then have no direct translation into $\mathcal{Q}_1$.

Once you see the basic structural problem, it is easy to find other instances. There is, for example, a certain family of views that go under the heading of **modal relativism**. The modal relativist about properties, for example, says that whether or not something has a property at a particular world is itself a relative matter. So for example, suppose that there is a particular drinking glass g at a possible world w that is filled with H_2O. Does that glass also have the property of being filled with water? The modal relativist about properties says that the answer depends on which world is actual. Relative to our world v being actual, g is filled with water at w, since our world is one in which H_2O plays the water role. But relative to a different world u in which some other substance XYZ plays the water role, g is not filled with water at w. This basic idea is easy to express using $\mathcal{E}$. We can just use a double-indexed *is filled with water* predicate:

$$\exists v(Av \wedge \exists w(Wgwv \wedge \exists u Wgwu))$$

But because this claim require two world variables, there is no way to express it in $\mathcal{Q}_1$.

---

can translate sentences that require two world variables only if the second world variable is only needed to express facts involving the accessibility relation. This sort of limited exception, though, will not help with the problematic cases.

Moreover, besides modal relativism about properties, there are other sorts of views you might want to hold: You might want to be a modal relativist about *ontology*. You might think that certain things exist at a world w from the perspective of one world v, but not from the perspective of some other world u.[5] Or you might be a modal relativist about *modality*. You might think that in certain cases, a world w is accessible from v from the perspective of x, but not accessible from v from the perspective of some other world y.[6]

The idea behind multi-dimensional modal logic is to build autonomous languages with fewer restrictions. In the next section, we are going to build a whole hierarchy of modal languages $\mathcal{Q}_n$, with each having the ability to directly translate claims that require n world variables. Most of the problematic cases require only two world variables, so can be expressed in $\mathcal{Q}_2$. For example:

$\Diamond\exists x{\downarrow}\forall y(x \neq y)$

$\Diamond Tas$

$\Diamond\otimes Tas$

$\Diamond{\uparrow}Tas$

$\Diamond(Wg \wedge \otimes\Diamond\otimes\neg Wg)$

The first sentence says that there could have been something that does not actually exist. The second says that Aristotle could have been taller than Socrates actually is. The third says that Aristotle could have been *shorter* that Socrates actually is. The fourth says that Aristotle could have been taller than Socrates. And the fifth says that there is a possible

5. See my (2019) for a defense of this sort of view.

6. See (Murray and Wilson 2012) and (Helle, Murray, and Wilson 2018) for more on modal relativism.

world at which a glass is filled with water, relative to the actual world, even though it is not filled with water relative to some other possible world.

Now of course, there are also other strategies for dealing with the expressive limitations of Q_1. You might, for example, try to use certain **indirect translations** in place of direct translations. In the first case, you might try expressing the idea that there could have been something that does not actually exist by quantifying over sets:

$$\exists x(\forall y(y \in x) \wedge \Diamond \exists z(z \notin x))$$

There is a set that has everything as a member and it could have been that something existed without being a member of that set. In the second case, you might try to express the idea that Socrates could have been taller than he actually is by quantifying over heights:

$$\exists x(Hax \wedge \Diamond \exists y(Hsy \wedge y > x))$$

There exists a height that Aristotle has and it could have been that there was a height Socrates had that was greater than the one Aristotle has. Given the right background assumptions, you might think, these represent perfectly good translations of the corresponding claims.

The problem with indirect translations is that they almost always come with additional commitments. In the first indirect translation, for example, we not only quantified over sets—we quantified over a set of *everything*. But while this might work in certain contexts, it will not work in others. Maybe there are contexts in which we want to deny that there are sets altogether. Or maybe there are contexts in which we are happy to accept that there are sets, but want to maintain that sets are wellfounded (and so deny that there is a set of everything). The advantage of direct translations is that we can remain neutral on these questions. There is no need to take any particular position on the nature of sets in order to accept that there could have been something that does not actually exist.

The second case is similar. Maybe there are contexts in which we can compare things across world by quantifying over things like heights. But this strategy also comes with certain additional commitments, and so will not work in all cases. For example, you might want to use comparisons across worlds to do science without numbers. But if the only way to express such comparison is by quantifying over things like masses and charges and distances, then no such thing is possible, since quantifying over masses and charges and distances means quantifying over abstract things, when quantifying over abstract things is precisely what we are trying to avoid. Direct translations, on the other hand, come with no such commitments, so represent a powerful tool for simplifying the metaphysical commitments of science. This is just the point we considered at length in chapter one.

Yet another advantage of direct translations is that they are completely systematic. So long as a sentence of a regimented language has the right form, there will be a translation into an appropriate multi-dimensional language. Indirect translations, on the other hand, are *ad hoc* by their very nature. By way of illustration, consider a temporal case. Using a temporal regimented language, we can make the following pair of claims:

$$\exists t(Pt \wedge \exists s(t > s \wedge T(otws)))$$

$$\exists t(Pt \wedge \exists s(t > s \wedge A(otws)))$$

The first says that Obama at the present time is taller than Washington at some time in the past. And the second says that Obama at the present time *admires* Washington at some time in the past. Using a temporal analogue of Q_1, we can perhaps express the first by quantifying over heights. But even if so, the same strategy will not work for the second sentence—and this despite the fact that they have exactly the same form. Quantification over things like heights, then, is at best a stopgap measure. It may help us express comparative relations across time, but is of no help whatsoever when it comes to *non*-comparative relations.

3.2. Propositional Languages

Now that we have some motivation for multi-dimensional languages, we can get to work actually building them. We are going to start by building a hierarchy of propositional languages in this section. Then we are going to build a corresponding hierarchy of multi-dimensional quantified languages in section 3.4. Along the way, we will also generalize the familiar notion of a Kripke model from one dimension to arbitrarily many.

Each n-dimensional modal language $\mathcal{P}_n$ extends the language of propositional logic with certain sentential operators. In particular, the basic syntax of each $\mathcal{P}_n$ includes:

Atomic Sentences:	$P, Q, R...$
Extensional operators:	$\wedge, \neg$
Intensional operators:	$\Box$ when $n \geq 1$
	$\uparrow$ when $n \geq 2$
	$\otimes_i$ for each $n \geq i \geq 2$
	Id when $n \geq 2$

The $\Box$, $\uparrow$, and $\otimes_i$ operators are one-place. The Id operator is zero-place, making it a kind of logical atomic sentence. The sentences of each $\mathcal{P}_n$ are then defined recursively in the way you would expect.[7] This gives us not just one, but an entire hierarchy of propositional modal languages. The most familiar are the language of propositional logic (which is $\mathcal{P}_0$) and the language of standard one-dimensional modal logic (which is $\mathcal{P}_1$). The rest of the languages are all multi-dimensional. Of those, the simplest is the two-dimensional language $\mathcal{P}_2$. As you can see from the above, that language extends propositional logic with exactly four operators, which are $\Box$, $\uparrow$, $\otimes$, and Id.[8]

At this point, there are different sorts of models we could use. To start, we are going to use the simplest, which are multi-dimensional **Leibniz models**. These are structures of the form:

$$\mathcal{M}_n = \langle W, [\![\cdot]\!] \rangle$$

The first element is a non-empty set of possible worlds. The second is a valuation function. The subscripted n indicates the dimensionality of the modal. Each language $\mathcal{P}_n$ is then modeled by the class of models with the corresponding number of dimensions. So $\mathcal{P}_1$ is modeled using the class of one-dimensional Leibniz models, $\mathcal{P}_2$ is modeled with the class of two-dimensional Leibniz models, and so on.

There are two things that make a Leibniz model multi-dimensional instead of just one-dimensional. The first is the valuation function. Every n-dimensional model comes with an

7. Every atomic sentence is a sentence. Moreover, if ϕ and ψ are sentences, then $\neg\phi$ and $\phi \wedge \psi$ and $\Box\phi$ and $\uparrow\phi$ and $\otimes_i\phi$ and Id are sentences. These are all the sentences, at least when n is finite. When n is infinite, we may also want to add infinitary formation rules that allow for things like infinite conjunctions and infinite strings of operators.

8. $\mathcal{P}_2$ has only one $\otimes_i$ operator, so we will generally drop the subscript, writing $\otimes$ instead of $\otimes_2$.

n-dimensional valuation function, which is a function that assigns each atomic sentence to a set of sequences of the appropriate length.[9] So:

$$[\![P]\!] \subset \mathbb{P}(W^n)$$

The second thing is that the central notion of truth, which is truth at a *sequence* of worlds rather than just truth at a world. This is defined recursively:

$$w_1w_2w_3... \models P \quad \text{iff} \quad w_1w_2w_3... \in [\![P]\!]$$

$$w_1w_2w_3... \models \neg\phi \quad \text{iff} \quad w_1w_2w_3... \not\models \phi$$

$$w_1w_2w_3... \models \phi \wedge \psi \quad \text{iff} \quad w_1w_2w_3... \models \phi \text{ and } w_1w_2w_3... \models \psi$$

$$w_1w_2w_3... \models \Box\phi \quad \text{iff} \quad uw_1w_2... \models \phi \text{ for every } u \in W$$

$$w_1w_2w_3... \models \uparrow\phi \quad \text{iff} \quad w_1w_1w_3... \models \phi$$

$$w_1...w_i... \models \otimes_i\phi \quad \text{iff} \quad w_i...w_1... \models \phi$$

$$w_1w_2w_3... \models \mathsf{Id} \quad \text{iff} \quad w_1 = w_2$$

As you can see, each operator corresponds to natural operation on sequences. The **box** operator $\Box$ pushes all the worlds in a sequence to the right, then puts a new world in the first position. When n is finite, the last world from the original sequence is deleted. The **up** operator $\uparrow$ copies the world from the first position into the second position. Each **swap** operator $\otimes_i$ swaps the world in the first position with the world in the i-th position. And finally, the **identity** operator Id checks to see if the first and second worlds are identical. A sentence is then true at a world w when true at the sequence $www...$ in which that world

9. Going forward, when I talk about sequences, I mean sequences of worlds of length n, where n is the dimensionality of the corresponding model.

occupies every position. And a sentence is true in a model when true at every world.

Notice that on the above way of thinking about things, we get the familiar one-dimensional Leibniz models as a special case. In particular, when $n = 1$, Leibniz models consist of a set of worlds and a valuation function assigning each atomic sentences to a set of worlds.[10] Truth at a world is then defined recursively in the usual way:

$$\begin{array}{lll} w \models P & \text{iff} & w \in [\![P]\!] \\ w \models \neg\phi & \text{iff} & w \not\models \phi \\ w \models \phi \wedge \psi & \text{iff} & w \models \phi \text{ and } w \models \psi \\ w \models \Box\phi & \text{iff} & u \models \phi \text{ for every } u \in W \end{array}$$

A sentence is then true in a model when true at every world.

Since the general case is a little abstract, it can be helpful to also consider the two-dimensional case. When $n = 2$, Leibniz models consists of a set of worlds and a valuation function. This valuation function assigns each atomic sentence to a set of *pairs* instead of just a set of worlds. The central notion of truth is then truth at a pair, which is defined recursively:

$$\begin{array}{lll} wv \models P & \text{iff} & wv \in [\![P]\!] \\ wv \models \neg\phi & \text{iff} & wv \not\models \phi \\ wv \models \phi \wedge \psi & \text{iff} & wv \models \phi \text{ and } wv \models \psi \\ wv \models \Box\phi & \text{iff} & uw \models \phi \text{ for every } u \in W \end{array}$$

10. Technically, the valuation assigns atomic sentences to sets of sequences of length one. But the sequences of length one are obviously isomorphic to the worlds. So we can, for practical purposes, think of the valuation function as assigning atomic sentences to sets of worlds.

$$wv \models \uparrow\phi \quad \text{iff} \quad ww \models \phi$$

$$wv \models \otimes\phi \quad \text{iff} \quad vw \models \phi$$

$$wv \models \mathsf{Id} \quad \text{iff} \quad w = v$$

Truth at a pair is a purely technical notion but, for heuristic purposes, you might think of it in any of the following ways:

ϕ is true at wv iff ϕ is true at w from the perspective of v

ϕ is true at wv iff ϕ is true at w relative to v

ϕ is true at wv iff ϕ is true at w on the assumption that v is actual

As noted earlier, each of our operators corresponds to a certain natural operations on pairs. The box operator pushes the first world into the second position, then fills the first position with an arbitrary world. The up operators copies the world from the first position into the second position. The swap operator exchanges the two worlds. And the identity operator checks to see if the two worlds are identical. A sentence is then true at a world w when true at the pair ww. And a sentence is true in a model when true at every world.

Now that we have our basic languages and models, it will be helpful to also have certain non-basic expressions. These include:

$$\Diamond\phi \quad \text{iff} \quad \neg\Box\neg\phi$$

$$\phi \vee \psi \quad \text{iff} \quad \neg(\neg\phi \wedge \neg\psi)$$

$$\phi \supset \psi \text{ ‘} \quad \text{iff} \quad \neg(\phi \wedge \neg\psi)$$

$$\phi \equiv \psi \quad \text{iff} \quad (\phi \supset \psi) \wedge (\psi \supset \phi)$$

$$\downarrow\phi \quad \text{iff} \quad \otimes\uparrow\phi$$

This gives all of our languages a possibility operator and the usual range of truth-functional connectives. They also get a second copy operator $\downarrow$ called **down**. Where up copies the world from the first position into the second position, down does the reverse. It copes the world from the second position into the first position.[11]

By this point, observant readers may have noticed that our basic operators are somewhat redundant. This because the up operator can be defined in terms of box and identity, given the following biconditional:

$$\uparrow\phi \equiv \Box(\mathsf{Id} \supset \phi)$$

This would seem to indicate that the up operator should simply be dropped from the list of basic operators. There are two reasons, though, that the up operator has been included. The first is that you might want to drop the identity operator from the language. You might, for example, want to deny that there is any sense to the question of whether or not there are distinct worlds that are also qualitative duplicates. Second, once we have a more general approach to models, the above biconditional can fail. So it is not true *in general* that up can be defined in terms of box and identity, even though this happens to be true when using Leibniz models.[12]

11. When $n > 2$, there are other sorts of operators we might want to define. For example, we might want to define a family of operators $\otimes_{ij}$ that swap worlds between arbitrary coordinates. Or we might want to define a family of operators $\uparrow_{ij}$ that copy worlds between arbitrary coordinates. Or we might want to define a family of operators Id_{ij} that test for identity between arbitrary coordinates. You get the idea. Working out how to define these operators is an interesting exercise.

12. More precisely: Once we get to Kripke models, up can be defined in terms of box and identity iff the accessibility relation is reflexive.

3.3. Multi-Dimensional Kripke Models

Leibniz models encode a certain view about the nature of possibility, one on which the range of possibilities is fixed as a matter of necessity. Whatever could have been possible is already possible. Whatever could have been necessary is already necessary. Objective modalities, like metaphysical necessity and physical necessity, are generally thought to satisfy these conditions, so this is perhaps not a decisive problem. But you might also find yourself in the minority, and so convinced that the range of possibilities is contingent. It would be nice, then, if we had models that let us make sense of this idea.

Even if you think that the range of objective possibilities is necessary, still, you might want to use modal logic to reason about other intensional notions like knowledge and permissibility. But in those cases, the assumptions embedded in Leibniz models will clearly fail. We do not always know what we know, nor are the actions we perform always permissible. But in order to make sense of the first idea, we will need to deny S4 and, in order to make sense of the second, we will need to deny T. Since Leibniz models validate both, this means that we will need a more general way to think about models. What's more, the reasons to adopt a multi-dimensional framework generally carry over from the modal case to these other sorts of notions. Just as you might want to express the idea that faculty meetings *could* have been less boring than they are, you might want to express the idea that they *should* be less boring than they are, or that you *thought* they would be less boring than they are. And these sorts of claims, as we have already seen, are naturally expressed using a multi-dimensional framework.

The natural way to generalize Leibniz models is by using multi-dimensional **Kripke**

models. These are just like Leibniz models, but with the addition of an accessibility relation $R \subset W^2$. This makes them structures of the form:

$$\mathcal{M}_n = \langle W, R, [\![\cdot]\!] \rangle$$

Truth at a sequence is defined the same way as before, with the exception of the clause for the box operator:

$$w_1 w_2 w_3 ... \models \Box\phi \quad \text{iff} \quad u ... w_i ... w_n \models \phi \text{ for every } u \in W \text{ such that } uw_1 \in R$$

In the one-dimensional case, this gives us:

$$w \models \Box\phi \quad \text{iff} \quad u \models \phi \text{ for every } u \in W \text{ such that } uw \in R$$

And in the two-dimensional case, we have:

$$wv \models \Box\phi \quad \text{iff} \quad uw \models \phi \text{ for every } u \in W \text{ such that } uw \in R$$

A sentence is then true at a world w when true at the sequence $www...$ in which that world occupies every position. And it is true in a model when true at every world.[13]

To my knowledge, multi-dimensional Kripke models appear nowhere in the literature.

13. An alternative approach would be to say that n dimensional models always have an $n+1$ place accessibility relation instead of an n place accessibility relation. So for example, in the two-dimensional case, we would use a *three*-place accessibility relation instead of a two-place relation. $\Box\phi$ would then be true at a pair wv iff there was a u such that $Ruwv$. This would let us make sense of the idea that whether or not a certain world w is accessible from v is itself a perspectival matter. From one world x, w is accessible from v. But from another world y, w is not accessible from v. Whether are any useful applications for this extra degree of relativity is an interesting question.

This is surprising, given their relative simplicity and flexibility. For whatever reason, the general focus has been on **concrete models**.[14] In the two-dimensional case, these are structures:

$$\mathcal{M}_2 = \langle S, [\![\cdot]\!] \rangle$$

The first element is a set of possible pairs. The second is a valuation function assigning each atomic sentence to a set of possible pairs, meaning that $[\![P]\!] \subset S$. Leibniz models are then the special case in which $P = W^2$. But while all Leibniz models are concrete models, not all concrete models are Leibniz models, since the set of possible pairs could also be a strict subset of W^2 meeting certain closure conditions. Those closure conditions are that $ww \in P$ and $vw \in P$ whenever $wv \in P$. These are needed to ensure that the up operator and swap operator are defined. A sentence is then true in a model when true at all possible pairs.

The problem with concrete models is that the set of possible pairs is in effect forced to play the role of both a set of worlds and an accessibility relation. You can see this by looking at the clause for the necessity operator, which is going to be something like:

$$wv \models \Box\phi \quad \text{iff} \quad uw \models \phi \text{ for every } uw \in S$$

This has certain unfortunate consequences. For example, sentences can only be true at a pair if that pair is in the set of possible pairs. So in order to make a sentence ϕ true at a world w, we have to include ww in the set of possible pairs. But once ww is in the set of possible pairs and ϕ is true at ww, it follows that $\Diamond\phi$. So concrete models validate the T

14. See Marx and Venema (1997) for more on this.

axiom. Moreover, they also validate a two-dimensional version of the B axiom, since the closure conditions force the set of possible pairs to be symmetric. We could try to fix the problem by weakening the closure conditions but, in that case, the swap operator will not always be defined.

Multi-dimensional Kripke models can avoid these problems because they have both a set of worlds and an accessibility relation, just like in the one-dimensional case. This gives them the flexibility we need to model weaker systems.

3.4. Quantified Languages

We built a hierarchy of propositional modal languages $\mathcal{P}_n$ in section 3.2. We are now going to build a corresponding hierarchy of quantified modal languages $\mathcal{Q}_n$, each of which have the following basic syntax:

Non-logical predicates:	$P, Q, R...$
Logical predicates:	$=$
Variables:	$x, y, z...$
Names:	$c, d, e...$
Quantifiers:	$\forall$
Extensional operators:	$\wedge, \neg$
Intensional operators:	$\Box$ when $n \geq 1$
	$\downarrow$ when $n \geq 2$
	$\otimes_i$ for each $n \geq i \geq 2$
	Id when $n \geq 2$

Every predicate is assigned a fixed finite arity and the sentences of each language are defined recursively in the usual way.[15] The non-basic expressions then include the ones from earlier, along with:

$\exists x\phi$ iff $\neg\forall x\neg\phi$.

Et iff $\exists x(t = x)$

This gives each of our languages both an existential quantifier and an existence predicate.

Once we have our languages, they can each be modeled using the corresponding class of quantified Kripke models. These are just like the propositional Kripke models from the last section, but have the extra structure needed to handle things like quantifiers and predicates.[16]

$$\mathcal{M}_n = \langle W, R, D, d, [\![\cdot]\!]\rangle$$

The first three elements are a domain of worlds, an accessibility relation, and a domain of possible individuals. The fourth is a domain function mapping each sequence of worlds to a set of possible individuals:

$$d : W^n \longrightarrow \mathbb{P}(D)$$

15. Let $t_1, ..., t_n$ be terms and P^n an arbitrary n-place predicate. Then $P^n(t_1, ..., t_n)$ and $(t_1 = t_2)$ are sentences. And if ϕ and ψ are sentences, then so are $\forall x(\phi)$, $(\phi \supset \psi)$, $\neg(\phi)$, $\Box(\phi)$, $\downarrow(\phi)$, $\otimes_i(\phi)$, and Id. Nothing else is a sentence. As in the propositional case, we may want to use infinitary formation rules when the number of dimensions is infinite. The assumption of fixed finite arities could also be easily dropped.

16. Generally speaking, we will refer to both propositional Kripke models and quantified Kripke models as just Kripke models, letting context determine which sort of models we have in mind.

Finally, we have an n-dimensional valuation function, which does two things:

$[\![c]\!] \in D$

$[\![P]\!] : W^n \longrightarrow \mathbb{P}(D^m)$

First, it assigns every name to a possible individual. And second, it assigns each m-place predicate to an n-dimensional intension, which is just a function from sequences of worlds (of length n) to sets of sequences of individuals (of length m). The central notion of truth is then truth at a sequence *relative* to a variable assignment. This is defined recursively, with the clauses for the sentential operators being the same as before. The new clauses are:

$w_1w_2w_3... \models_\sigma P(t_1, ..., t_n)$ iff $\sigma(t_1)...\sigma(t_m) \in [\![P]\!](w_1w_2w_3...)$

$w_1w_2w_3... \models_\sigma t_1{=}t_2$ iff $\sigma(t_1){=}\sigma(t_2)$

$w_1w_2w_3... \models_\sigma \forall x\phi(x)$ iff every τ is such that $w_1w_2w_3... \models_\tau \phi(x)$

A sentence is true at a sequence when true at that sequence relative to all variable assignments. A sentence is true at a world when true at the sequences in which that world occupies every position. And a sentence is true in a model when true at every world.[17]

As before, this gives us familiar one-dimensional Kripke models as the special case in which $n = 1$. The two-dimensional case in which $n = 2$ is then as follows: First we have a set of worlds, an accessibility relation, and a domain of possible individuals. This is all the same as in the one-dimensional case. What makes two-dimensional models different is

17. A variable assignment σ is a function mapping each term in the language to a member of D. In the case of names, it does so in a way that matches the valuation function. So $\sigma(c) = [\![c]\!]$. In the above recursive clauses, the variable assignment τ is just like σ, with the possible exception that $\tau(x) \neq \sigma(x)$ and the added requirement that $\tau(x) \in d(w_1w_2w_3...)$.

then two things. First, the domain function maps *pairs* of worlds to sets of individuals. And second, the valuation function maps every predicate to a function mapping *pairs* of worlds to sets of sequences of individuals of the appropriate length.

What does this all mean exactly? You might think of $a \in d(wv)$ as encoding the fact that a exists at w from the perspective of v. Our two-dimensional models thus allow for a kind of modal ontological relativity, since we can have $a \in d(wv)$ but not $a \in d(wu)$. The individual a may exist at w from the perspective of v, but not exist at w from the perspective of some other world u.[18] Once we get to the valuation function, one-place predicates are mapped to functions from pairs of worlds to sets of individuals. Two-place predicates are mapped to a function from pairs of worlds to sets of pairs of individuals. And so on. How you should think about these functions depends somewhat on the intended reading of the relevant predicates. But for example, if W denotes the *being filled with water* property, you might think of $a \in [\![W]\!](wv)$ as indicating that a has the property of being filled with water at w from the perspective of v. And if T denotes the *taller than* relation, you might think of $ab \in [\![T]\!](wv)$ as indicating that a at w is taller than b at v.

Once we get to truth in a model, the important notion is truth at a pair relative to a variable assignment. This is defined recursively, with the clauses for the sentential operators being the same as before. The new clauses are:

18. When this relativity is unwanted, we can just use the class of frames with non-relative domains. That is, we can just use those frames in which $d(wv) = d(wu)$ for all worlds.

$$wv \models_\sigma P(t_1, ..., t_m) \quad \text{iff} \quad \sigma(t_1)...\sigma(t_n) \in [\![P]\!](wv)$$

$$wv \models_\sigma t_1{=}t_2 \quad \text{iff} \quad \sigma(t_1){=}\sigma(t_2)$$

$$wv \models_\sigma \forall x\phi(x) \quad \text{iff} \quad \text{every } \tau \text{ is such that } wv \models_\tau \phi(x)$$

A sentence is true at ww when true at ww relative to all variable assignments σ. A sentence is true at a world w when true at the pair ww. And a sentence is true in a model when true at all worlds.

3.5. Proof Theory for PML

Proof systems for quantified modal logic are naturally built in stages. First, we build an appropriate propositional modal logic. Then we add the rules and axioms of an appropriate quantified predicate logic. Finally, we glue everything together by adding certain further principles as needed. In what follows, we are going to use this process to build various two-dimensional quantified modal logics. We are focusing on the two-dimensional case for simplicity and concreteness. But the basic approach extends to higher dimensions as well.

The propositional modal logic $\mathbf{K} = \mathbf{K_1}$ is the foundation of one-dimensional quantified modal logic. The first step towards building a multi-dimensional QML, then, is to construct an appropriate multi-dimensional analogue. Since our ultimate goal is to build a two-dimensional QML, that means building $\mathbf{K_2}$.

The principles of $\mathbf{K_2}$ fall into two broad categories. The first category consists of those principles that are straightforward analogues of the rules and axioms of $\mathbf{K_1}$.

MP $\phi \supset \psi, \phi \Rightarrow \psi$

RN $\phi \Rightarrow \Box\phi$

$\phi \Rightarrow \uparrow\phi$

$\phi \Rightarrow \otimes\phi$

K $\Box(\phi \supset \psi) \supset (\Box\phi \supset \Box\psi)$

$\uparrow(\phi \supset \psi) \supset (\uparrow\phi \supset \uparrow\psi)$

$\otimes(\phi \supset \psi) \supset (\otimes\phi \supset \otimes\psi)$

PL ϕ when ϕ is a theorem of propositional logic

As you can see, the system has two basic rules, which are modus ponens and the rule of necessitation. The axiom K then says that all three operators distribute over material conditionals. And PL says that the theorems of classical propositional logic are also theorems of $\mathbf{K_2}$. These together ensure that all of the relevant operators are normal.

The second category of principles consists of axioms governing the interaction of our various operators. These are what give $\mathbf{K_2}$ its two-dimensional character.

A1 $\uparrow\neg\phi \equiv \neg\uparrow\phi$

A2 $\otimes\neg\phi \equiv \neg\otimes\phi$

A3 $\uparrow\otimes\phi \equiv \uparrow\phi$

A4 $\otimes\otimes\phi \equiv \phi$

A5 $\uparrow\phi \supset \Box\otimes\uparrow\phi$

A6 $\uparrow\Box\phi \equiv \Box\phi$

A7 $\mathsf{Id} \supset (\uparrow\phi \equiv \phi)$

A8 $\uparrow\mathsf{Id}$

A9 Id

Thinking in terms of pairs, A1 says that the result of copying is always unique. A2 says that the result of swapping is always unique. A3 says that copying and then swapping is the same as just copying. A4 says that swapping twice is the same as not swapping at all. A5 says that the box operator has no effect on the second coordinate. A6 says that accessibility is fully determined by the first coordinate. A7 says that when two worlds are identical, copying does nothing. A8 says that copying always results in identical worlds. And finally, A9 says that our point of departure in modal space is always a *world*, rather than just an arbitrary pair. Note that in order to avoid triviality, the rule of necessitation needs to be restricted to exclude this last axiom.[19] If the identity operator is dropped from the language, then the last three axioms are replaced with the following two:

A7* $\uparrow(\uparrow\phi \equiv \phi)$

A8* $\uparrow\phi \equiv \phi$

The rule of necessitation is then restricted to exclude A8* instead of A9.

Once we have $\mathbf{K_2}$, stronger systems can be build by adding various characteristic axioms. This process is familiar from the one-dimensional case, but the axioms themselves are somewhat different. Some of the most important are listed in the table below. As you can see, all of the usual one-dimensional axioms have two-dimensional analogues. An axiom that will play an especially prominent role later on is TB, which is so-called because it is equivalent to the conjunction of T and B.

19. That is, RN says that $\Box\phi$, $\downarrow\phi$, and $\otimes$ are theorems when ϕ can be proved without A9.

	1D Axiom	2D Axiom	Frames
D	$\Box\phi \supset \Diamond\phi$	$\Box\phi \supset \Diamond\phi$	serial
T	$\Box\phi \supset \phi$	$\Box\phi \supset \uparrow\phi$	reflexive
TB		$\Box\phi \supset \otimes\phi$	reflexive and symmetric
B	$\Diamond\Box\phi \supset \phi$	$\Diamond\Box\phi \supset \Diamond\otimes\phi$	symmetric
S4	$\Box\phi \supset \Box\Box\phi$	$\Box\uparrow\phi \supset \Box\Box\uparrow\phi$	transitive
S5	$\Diamond\Box\phi \supset \Box\phi$	$\Diamond\Box\phi \supset \Box\Box\otimes\phi$	euclidean

As far as I can tell, there is no (non-conjunctive) analogue in the one-dimensional case, so that space in the left column has been left blank.

3.6. Proof Theory for QPL

Once we have a basic two-dimensional PML, the next step towards building a two-dimensional QML is identifying an appropriate quantified predicate logic. For our purposes, that quantified predicate logic is going to be the free logic **F**.

UG $\phi \Rightarrow \forall x\phi$

B1 $\forall x\phi x \supset (Et \supset \phi t)$

B2 $\forall x(\phi \supset \psi) \supset (\forall x\phi \supset \forall x\psi)$

B3 $\phi \supset \forall x\phi$, when x is not free in ϕ

B4 $t = t$

B5 $t = s \supset (\phi \supset \phi[t/s])$

F $\forall xEx$

First, we have the rule of universal generalization. There are then various axioms governing the universal quantifier and various axioms governing identity.[20] We then have the axiom F, which is so-called because it is what distinguished **F** from classical predicate logic, which has the stronger axiom:

C $\quad$ *Et*

When combined with the rule of necessitation, classical predicate logic entails that everything exists necessarily. But in certain contexts, this is a result we would like to avoid. So we are using the free logic **F** instead.[21]

3.7. Proof Theory for QML

You might have hoped that we could build a sensible quantified modal logic just by adding **F** to $\mathbf{K_n}$. But in fact, this is not the case. The combined logic $\mathbf{FK_1}$ is not complete with respect to *any* class of Kripke frames. And the same goes for $\mathbf{FK_2}$. What this means is that in order to build a complete system, further principles need to be added.[22]

Spelling this out a bit, a QML is complete only if it passes two important tests. The first is whether the system can prove certain modal principles involving identity. These include the **preservation of identity** and the **preservation of distinctness**.

20. For axiom B5, $\phi[t/s]$ is a sentence that result from ϕ by replacing any number of instances of s with t.

21. Note that besides the above, **F** also includes MP and PL, which are not listed here, since they were listed earlier.

22. Going forward, when we say that a QML is complete, we mean complete with respect to some class of Kripke models.

PI $\quad t = s \supset \Box(t = s)$

$t = s \supset \uparrow(t = s)$

$t = s \supset \otimes(t = s)$

PD $\quad t \neq s \supset \Box(t \neq s)$

$t \neq s \supset \uparrow(t \neq s)$

$t \neq s \supset \otimes(t \neq s)$

The preservation of identity can be proved using the fact that all of the relevant operators are normal. The situation is somewhat different, though, when we get to the preservation of distinctness. $\text{PD}_{\otimes}$ can be proved using $\mathbf{FK_2}$. We can also show that $\text{PD}_{\uparrow}$ entails $\text{PD}_{\Box}$. The problem is that $\mathbf{FK_2}$ and $\text{PD}_{\uparrow}$ are independent. So there is no way to prove the full preservation of distinctness in $\mathbf{FK_2}$.[23]

The second test is whether a system allows us to generalize in a sufficiently wide range of contexts. In ordinary quantified predicate logic, for example, we can generalize in the scope of material conditionals. This point might be put more precisely by saying that the rule of **conditional universal generalization** is valid:

CUG $\quad (\phi \supset \psi) \Rightarrow (\phi \supset \forall x\psi)$ when x is not free in $(\phi \supset \forall x\psi)$

Some systems treat this as a basic rule. Others treat it as derived. Once we move to quantified modal logic, a system is complete only if we can do the same with **strict conditionals**. These are sentences $\phi \dashv\!\!3\, \psi$ of the form

23. The one-dimensional results described in this section are all well-known. The two-dimensional results will be proved later on.

$\phi_1 \supset (\phi_2 \supset \ldots (\phi_n \supset \psi) \ldots)$

where any of the embedded material conditionals are prefaced by any number of intensional operators. So for example, all of the following count as strict conditionals:

$\Box(\phi \supset \psi)$

$\otimes(\phi_1 \supset (\phi_2 \supset \psi))$

$\phi_1 \supset \Box(\phi_2 \supset \otimes\!\downarrow\!(\phi_3 \supset \psi))$

What we can show is that a system is complete only if it validates **modal universal generalization**, which says precisely that we can generalize in the scope of strict conditionals.

MUG $\quad (\phi \dashv\!3\ \psi) \Rightarrow (\phi \dashv\!3\ \forall x\psi)$ when x is not free in $(\phi \dashv\!3\ \forall x\psi)$

The problem is that systems like $\mathbf{FK_1}$ and $\mathbf{FK_2}$ do not validate MUG. And so it follows that the systems are not complete.

There are two main approaches to fixing the problem. The first is to take certain mixed principles—like MUG or PD—as basic. These principles are "mixed" in the sense that they directly govern the interaction of modal ideology and quantificational ideology. Using this strategy, we can build both $\mathbf{Q_1}$ and $\mathbf{Q_2}$. The first is the logic of all one-dimensional Kripke models, and the second is the logic of all two-dimensional Kripke models.

System	**Components**	**Frames**
$\mathbf{Q_1}$	$\mathbf{FK_1}$, MUG, $\text{PD}_\Box$	all
$\mathbf{Q_2}$	$\mathbf{FK_2}$, MUG, $\text{PD}_\uparrow$	all
$\mathbf{QB_1}$	$\mathbf{FK_1}$, B	symmetric
$\mathbf{QTB_2}$	$\mathbf{FK_2}$, TB	reflexive and symmetric

The other strategy is to strengthen the underlying propositional modal logic. In the one-dimensional case, we can derive both MUG and PD in any system that has the B axiom. This means that, for example, the system $\mathbf{QB_1}$ is complete with the respect to the class of all symmetric Kripke frames. Later, we are going to show that in the two-dimensional case, we can derive MUG and PD in any system that has TB. This means that the system $\mathbf{QTB_2}$, for example, is complete with respect to the class of reflexive and symmetric Kripke frames.

Both of these strategies strike me as completely sensible. The advantage of the first is flexibility. If you are doing quantified epistemic logic, say, you will probably want to deny that if you might know a proposition, then that proposition is true. But in that case, the B axiom will fail, so the first strategy is the only game in town. The advantage of the second strategy is simplicity. If you are in a context where you have B axiom (in the one-dimensional case) or TB axiom (in the two-dimensional case), you might as well use them to derive principles like MUG and PD. The main point for our purposes is that *both* strategies are still available after moving from one-dimensional QML to two-dimensional QML.

3.8. Barcan Formulas

Once we have a complete system, stronger systems can be built by adding characteristic axioms. We saw some of these when we axiomatized two-dimensional PML in section 3.5. Once we move to a two-dimensional QML, the main new possibilities are axioms governing the interaction of operators and quantifiers.

	Axiom	Frames
BF	$\forall x \Box \phi \supset \Box \forall x \phi$	$\Box$-decreasing
	$\forall x \uparrow \phi \supset \uparrow \forall x \phi$	$\uparrow$-decreasing
CBF	$\Box \forall x \phi \supset \forall x \Box \phi$	$\Box$-increasing
	$\uparrow \forall x \phi \supset \forall x \uparrow \phi$	$\uparrow$-increasing

BF and CBF are familiar in the case of the necessity operator. The first tells us that everything that could have existed already exists. And the second says that everything that exists also exists necessarily. Putting them together, we get the claim that **ontology is fixed**. The things that exist do not change as we switch from considering one possible world to considering another. BF and CBF are somewhat less familiar in the case of the up operator. The first says that whatever exists at a world from its own perspective also exists at that world from the perspective of any other world. The second says that whatever exists at a world from the perspective of any world exists at that world from its own perspective. Putting these together, we get the idea that **ontology is non-relative**. What exists at a world does not depend on which world is actual.

Putting the issue more precisely, once we have a two-dimensional QML, we can distinguish two senses in which a frame might be increasing or decreasing.

$\Box$-increasing	$d(wv) \subset d(uw)$ when Ruw
$\uparrow$-increasing	$d(wv) \subset d(ww)$
$\Box$-decreasing	$d(uw) \subset d(wv)$ when Ruw
$\uparrow$-decreasing	$d(ww) \subset d(wv)$

Whether a domain is $\Box$-increasing or $\Box$-decreasing is a matter of whether the domain always grows or shrinks as we consider *different* possible worlds from the perspective of the *same* world. Whether a domain is $\uparrow$-increasing or $\uparrow$-decreasing, on the other hand, is a matter of whether the domain grows or shrinks as we consider the *same* possible world from the perspective of *different* worlds. Having these two senses in which a domain might be increasing or decreasing thus raises the possibility of holding certain novel—and for the most part unexplored—positions in modal ontology.[24]

3.9. PML Completeness

We now know how to build a variety of two-dimensional system. Next, we are going to show that they are complete. First, we are going to show that the propositional modal logic $\mathbf{K_2}$ is complete. Then we are going to show that the quantified modal logics $\mathbf{Q_2}$ and $\mathbf{QTB_2}$ are complete.

The most familiar strategy for proving completeness is to build a canonical model. A canonical model is one in which the domain of worlds is the set of all maximal consistent sets of sentences. The valuation function then assigns each atomic sentences to the set of

24. For example, you might hold that ontology is both fixed and relative. On this view, from the perspective of any world v, the same things exist at every possible world w, and so ontology is a completely necessary matter from the perspective of every world. But nevertheless, it may be that *different things* exist at every world from the perspective of v than exist at every world from the perspective of some other world u. Whether this is a view you should hold is a substantive question in metaphysics, so a matter best left for another time. For a defense of this sort of relative necessitism, see (Murray and Wilson 2012). For my own part, I am inclined toward relative contingentism, so am inclined to deny BF and CBF in both forms.

worlds that have that sentence as a member. And the accessibility relation is similarly fixed by the facts about which sentences are members of which sets.[25]

The problem with the canonical approach is there is no straightforward way of extending to two or more dimensions. In the two-dimensional case we can, of course, form the set of all maximal consistent sets of sentences, and so we might try using that set as our domain of worlds. But then how are we going to build the valuation function? In the one-dimensional case, we can just stipulate that a sentence is true at a world iff it is a member of that world. But the equivalent move in the two-dimensional case—saying that a sentence is true at a pair iff is a member of that pair—makes no sense. Sentences are not members of pairs of worlds. We could try identifying truth at a pair with, say, being a member of the world in the first coordinate. But this only works if the same sentences are always true at pairs *wv* and *wu*, which is not generally the case. Other ideas along these lines will fair no better, and for similar reasons.

The solution is to use what Blackburn, Rijke, and Venema (2010) call the **step-by-step approach**. The step-by-step approach, unlike the canonical approach, builds models by adding worlds one at a time. In the one-dimensional case, we might start by adding a single world. We then assign that world to a maximal consistent set of sentences. Unlike the canonical approach, then, we are not *identifying* worlds with maximal consistent sets. We are associating them. At that point, what we have is *almost* a model, but will almost certainly have certain gaps. So we make a list of the gaps and fill them in by adding more worlds one-by-one. Once all the gaps are filled, we have a model, and so have proved

25. For those unfamiliar with canonical models, see chapter six of (Hughes and Cresswell 1996).

completeness. As you will see in what follows, this basic process can be naturally extending to two dimensions and beyond.

Going forward, we will need to distinguish between two senses in which a set of sentences can be consists. We will say that a set is **strongly consistent** when no contradiction can be derived in $\mathbf{K_2}$. And we will say that a set is **weakly consistent** when no contradiction can be derived in $\mathbf{K_2}$ without using A9. Generally speaking, when we talk about a set being consistent full stop, what we mean is that it is weakly consistent. To simplify matters, we will also assume throughout that all of our languages are countable, though this is also an assumption that could be easily dropped. Without further ado then, let's get started.

Definition 3.9.1. *A **network** is a tuple $\langle W, R, f\rangle$ consisting of a set of worlds W, an accessibility relation $R \subset W^2$, and a partial function f assigning elements of W^2 to maximal consistent sets.*

Definition 3.9.2. *A network is **coherent** when:*

(C1) $\otimes\phi \in f(wv)$ *iff* $\phi \in f(vw)$

(C2) $\uparrow\phi \in f(wv)$ *iff* $\phi \in f(ww)$

(C3) $\mathsf{Id} \in f(wv)$ *iff* $w = v$

(C4) *If* $wv \in R$, *then* $f(wv)$ *is defined.*

(C5) *If* $\phi \in f(wv)$ *and* $uw \in R$ *then* $\Diamond\phi \in f(uw)$

Definition 3.9.3. *A network $\mathcal{N}$ is **saturated** when:*

If $\Diamond\phi \in f(wv)$, *then there is a* $uw \in R$ *such that* $\phi \in f(uw)$

Definition 3.9.4. *A network is **perfect** when it is both coherent and saturated.*

Definition 3.9.5. *A network $\mathcal{N}$ and a model $\mathcal{M}$* ***correspond*** *if they have the same frame and*

$$\mathcal{M}, wv \models \phi \text{ iff } \phi \in f(wv)$$

whenever $f(wv)$ is defined.

Readers familiar with the step-by-step approach may have noticed that we are using *partial* networks, which is somewhat non-standard. That is, we are allowing for the idea that a network may contain pairs of worlds wv such that $f(wv)$ is not defined. Moreover, not only are we allowing networks to be partial, we are even allowing *perfect* networks to be partial. This means, in effect, that even a perfect network may not fully determine which sentences are true at which worlds.

Now, you might think that this should all lead to disaster. Surely the point of building a perfect network is to fix on a unique corresponding model? And so surely we need perfect networks to be *total* instead of partial? But here is why we are using partial networks—and why this can be expected to work. We said earlier that a sentence is true in a model iff it is true at all worlds. That is, it is true in a model iff it is true at all pairs ww with the same world in each coordinate. The **realistic** pairs are then those pairs wv such that either $w = v$ or Rwv or Rvw. What we could show, with a bit of work, is that truth in a model is full determined by the realistic pairs. The unrealistic pairs are irrelevant. What a perfect network does, then, is assign a maximal consistent set to every *realistic* pair. Some of the unrealistic pairs may be left out. A perfect network can then have many corresponding models because there are many different ways of arbitrarily filling in truth at the undefined—and therefore unrealistic—pairs. In the next lemma, for example, we are going to construct a corresponding model by arbitrarily choosing to make all atomic sentences false at all undefined pairs. But other arbitrary choices would work just as well.

This explains why using partial networks can be expected to work. But why is it necessary? The use of partial networks is necessary because there can be cases in which no maximal consistent set can be assigned to an unrealistic pair without making the network incoherent. Using partial networks, then, gives us a simple way of sidestepping the problem.

Lemma 3.9.1 (Truth Lemma)**.** *Every perfect network $\mathcal{N}$ has a corresponding model $\mathcal{M}$.*

Proof. Let $\mathcal{N}$ be a perfect network. We then construct a corresponding model $\mathcal{M}$ by using the same frame, with the valuation function set to

$$[\![\phi]\!] = \{\langle w, v\rangle \mid f(wv) \text{ is defined and } \phi \in f(wv)\}$$

for every atomic sentence ϕ. That the resulting model is in fact a *corresponding* model is then shown using a straightforward induction on the complexity of sentences. ∎

Lemma 3.9.2. *Every strongly consistent set has a countable coherent network. In particular, given any any strongly consistent set S, there is a countable coherent network $\mathcal{N} = \langle W, R, f\rangle$ such that $S \subset f(wv)$ for some $w, v \in W$.*

Proof. Let S be any strongly consistent set. There is thus a maximal consistent set T extending S by Lindenbaum's lemma. There are then two cases. If $\mathsf{Id} \in T$, then we construct an $\mathcal{N}$ such that $W = \{w\}$ and $R = \emptyset$ and $f(ww) = T$. On the other hand, if $\mathsf{Id} \notin T$, we construct an $\mathcal{N}$ such that:

$$W = \{w, v\}$$

$$R = \emptyset$$

$$f(wv) = T$$

$$f(vw) = \{\phi \mid \otimes\phi \in T\}$$

$$f(ww) = \{\phi \mid \uparrow\phi \in T\}$$

$$f(vv) = \{\phi \mid \otimes\uparrow\phi \in T\}$$

Given the axioms, it can be easily verify that in either case, the result is a coherent network of the required kind. ■

We now know that every strongly consistent set of sentences has a coherent network. But for all we have said, that network may have certain defects that prevent it from being saturated. We are now going to show how to fix those defects, with the main step being the proof of something called the repair lemma.

Definition 3.9.6. *Let $\mathcal{N} = \langle W, R, f\rangle$ be any network. A **defect** is a triple $\langle w, v, \Diamond\phi\rangle$ such that $\Diamond\phi \in f(wv)$, but there is no $u \in W$ such that $\langle u, w\rangle \in R$ and $\phi \in f(uw)$.*

Definition 3.9.7. *A network $\mathcal{N}^*$ **extends** the network $\mathcal{N}$ when $W \subset W^*$ and $R \subset R^*$ and $f(wv) = f^*(wv)$ whenever $f(wv)$ is defined.*

Lemma 3.9.3. *If S is consistent and $\Diamond\phi \in S$, then there is a maximal consistent set T extending $S^* = \{\psi \mid \Box\psi \in S\} \cup \{\phi\}$.*

Proof. Suppose that S is consistent, but S^* not consistent. So there are $\Box\psi_1, ..., \Box\psi_n \in S$ such that $\psi_1 \wedge ... \wedge \psi_n \supset \neg\phi$. We thus have $\Box\psi_1 \wedge ... \wedge \Box\psi_n \supset \Box\neg\phi$ by RN and K. But then $\Box\neg\phi \in S$, meaning that S is not consistent, which is contrary to assumption. So S^* is consistent. There is thus a maximal consistent set T extending S^* by Lindenbaum's lemma. ■

Lemma 3.9.4 (Repair Lemma)**.** *For any defect of any countable coherent network $\mathcal{N}$, there is a countable coherent network $\mathcal{N}$ extending $\mathcal{N}^*$ without that defect.*

Proof. Let $\mathcal{N}$ be any coherent network with the flaw $\langle w, v, \Diamond\phi\rangle$. Furthermore, let $S = \{\psi \mid \Box\psi \in f(wv) \cup \{\phi\}$. We then check to see if there is already a $u \in W$ such that $S \subset f(uw)$. If so, then all we have to do is let $W^* = W$ and $R^* = R \cup \{\langle u, w\rangle\}$ and $f^* = f$. On the other hand, suppose that there is no such $u \in W$. In that case, we have to add one. So we choose any $u \notin W$ and let $W^* = W \cup \{u\}$ and $R^* = R \cup \{u, w\}$. We then let:

$$\begin{aligned} f(wv) &= \text{an arbitrary maximal consistent set } T \text{ extending } S \\ f(vw) &= \{\phi \mid \otimes\phi \in T\} \\ f(ww) &= \{\phi \mid \uparrow\phi \in T\} \\ f(vv) &= \{\phi \mid \otimes\uparrow\phi \in T\} \end{aligned}$$

That there is such a maximal consistent set follows from the previous lemma. Using the axioms, it can then be confirmed that in either case, we have a coherent network $\mathcal{N}^*$ extending $\mathcal{N}$, but without the original flaw. ■

Lemma 3.9.5. *Every countable coherent network $\mathcal{N}$ can be extended to a countable perfect network $\mathcal{N}^*$.*

Proof. Let $\mathcal{N}$ be a countable coherent network and W^* the result of adding countably many new worlds to W. The set of potential flaws is then $W^* \times W^* \times \mathsf{Sent}$, which is also countable. This means that we can form an enumeration of the potential flaws, which we then use to inductively define a chain of networks. For the base case, $\mathcal{N}_0 = N$. Now suppose that we have a network N_n and want to construct N_{n+1}. First, we check to see if the potential flaw $n + 1$ is a flaw of N_n. If not, then we move on, letting $N_{n+1} = N_n$. If so, then we repair the flaw using the method from the repair lemma, drawing unused worlds from W^* as needed. The result of the repair is then N_{n+1}. We then claim that the union N^* of this chain of networks is a perfect network extending $\mathcal{N}$.

$\mathcal{N}^*$ clearly extends $\mathcal{N}$. That $\mathcal{N}^*$ is coherent follows from the fact that (a) the base case is coherent, (b) the inductive step preserves coherence, and (c) the taking of unions preserves coherence. To show that $\mathcal{N}^*$ is saturated, suppose for reductio that it has a certain flaw $\langle w, v, \Diamond\phi\rangle$. Every flaw is a potential flaw, so thus has a place in our ordering. This means that we can refer to the alleged flaw as n. But then by construction, there is a network $\mathcal{N}_n$ in our chain that does not have n as a flaw. And if $\mathcal{N}_n$ does not have n as a flaw, then neither does any network extending $\mathcal{N}_n$. So n is not a flaw of $\mathcal{N}^*$ either, contrary to assumption. This means that $\mathcal{N}^*$ has no flaws and is therefore perfect. ∎

Theorem 3.9.1 (Completeness). *Every strongly* $\mathbf{K_2}$*-consistent set of sentences has a two-dimensional propositional Kripke model.*

Proof. Immediate from the preceding. ∎

Now that we have a completeness result for $\mathbf{K_2}$, you might also want similar results for stronger logics. For example, you might want to show that the logic $\mathbf{KTB_2}$ is complete with respect to the class of reflexive and symmetric two-dimensional Kripke frames. That proof is basically the same as the one we just gave. The only difference is that we would have to strengthen the definition of coherence to include the reflexivity and symmetry of the accessibility relation. At every stage in the process, we would then have to confirm that these additional coherence requirements were met.

3.10. QML Completeness

Now that we have a completeness result for $\mathbf{K_2}$, we would like to prove a similar result for $\mathbf{Q_2}$. This can be done using the **method of possible names**.[26] The basic idea is to replace maximal consistent sets with sets that also meet a certain further condition. Roughly speaking, that further condition is that whenever the set says that there *could* have been a possible individual, that possible individual also has a name. The difference between this method and what you might call the **method of actual names** is that we are using free logic instead of a classical logic. And so from the fact that a possible individual has a name, it need not follow that the possible individual in fact exists.

Proving that $\mathbf{Q_2}$ is complete is structurally similar to proving that $\mathbf{K_2}$ is complete, with a few key modifications. For this reason, we are not going to go through the proof in anything like exhaustive detail. Instead, we are going to indicate points at which the completeness proof for $\mathbf{K_2}$ needs to be modified and show how those modifications can be made. This will all be familiar to those have used the method of possible names in the one-dimensional case.

Definition 3.10.1. *A set S is **strictly omega complete** when:*

> *If $\phi \dashv (Ec \supset \psi c) \in S$ for every constant c that does not appear in $\phi \dashv \forall x \psi x$, then $\phi \dashv \forall x \psi x \in S$*

Definition 3.10.2. *A set of sentences is **superb** when it is maximal, consistent, and strictly omega complete.*

26. See Garson (2001) and chapter sixteen of Hughes and Cresswell (1996) for more on this.

Once we have the notion of a superb set, the definition of a network is shifted accordingly. That is, instead of mapping pairs of worlds to maximal consistent sets, a network maps pairs of worlds to superb sets. The definitions of coherence and satisfaction and correspondence remain the same. What we need to do then is replace the old version of the truth lemma with a new one, since we now have quantified Kripke models instead of propositional Kripke models.

Lemma 3.10.1 (Truth Lemma). *Every perfect network $\mathcal{N}$ has a corresponding Kripke model $\mathcal{M}$.*

Proof. Let $\mathcal{N} = \langle W, R, f \rangle$ be a perfect network assigning pairs of worlds to sets of sentences of $\mathcal{L}$. We then construct a corresponding model $\mathcal{M} = \langle W, R, D, d, [\![\cdot]\!] \rangle$ as follows. First, we define certain sets of terms:

$$\bar{c} = \{d \mid (c = d) \in f(wv) \text{ for any } w, v \in W\}$$

We then use these sets to construct $\mathcal{M}$:

$$\begin{aligned} D &= \{\bar{c} \mid c \in \mathcal{L}\} \\ d(wv) &= \{\bar{c} \mid Ec \in f(wv)\} \\ [\![c]\!] &= \bar{c} \\ [\![P]\!](wv) &= \{\langle \bar{c}_1, ..., \bar{c}_n \rangle \mid P(c_1, ..., c_n) \in f(wv)\} \end{aligned}$$

The result is that $\mathcal{M}$ is clearly a model. That it it corresponds to $\mathcal{N}$ can then be shown using an induction on the complexity of sentences. The main observation is that this requires both the preservation of identity and the preservation of distinctness, both of which are valid in $\mathbf{Q_2}$. ■

Once we have the truth lemma, proving completeness is reduced to showing that every strongly consistent set has a perfect network. The first step in that process is proving following lemma, which is a kind of replacement for Lindenbaum's lemma.

Lemma 3.10.2. *Every consistent set of sentences S in language $\mathcal{L}$ can be extended to a superb set T in an extended language $\mathcal{L}^+$.*

Proof. Suppose that S is a consistent set of sentences in $\mathcal{L}$. We then extend the language $\mathcal{L}$ to the language $\mathcal{L}^+$ by adding countably many new constants. Since $\mathcal{L}^+$ has countably many sentences, we can fix an enumeration of those sentences and use it to build the following chain:

(1) $T_0 = S$

(2) If δ_{n+1} is of the form $\neg(\phi \strictif \forall x\psi x)$ and $T_n \cup \{\neg(\phi \strictif \forall x\psi x)\}$ is consistent, then $T_{n+1} = T_n \cup \{\neg(\phi \strictif \forall x\psi x), \neg(\phi \strictif (Ec \supset \psi c))\}$, where c is a constant that has not yet appeared in the chain.

(3) If δ_{n+1} is not of the form $\neg(\phi \strictif \forall x\psi x)$ and $T_n \cup \{\phi\}$ is consistent, then $T_{n+1} = T_n \cup \{\phi\}$.

(4) If δ_{n+1} is not of the form $\neg(\phi \strictif \forall x\psi x)$ and $T_n \cup \{\phi\}$ is inconsistent, then $T_{n+1} = T_n \cup \{\neg\phi\}$.

(5) $T = \bigcup T_i$

We then claim that T is a superb set extending S. To show this, we need to show that S is a subset of T, that T is maximal, that T is consistent, and that T is strictly omega complete. The interesting case is showing that the first recursive clause preserves consistency. To that end, suppose that $T_n \cup \{\neg(\phi \strictif \forall x\psi x)\}$ is consistent and let c be any constant not appearing in that set. Now suppose for reductio that

$$T_n \cup \{\neg(\phi \mathrel{-\!\!3} \forall x \psi x)\} \cup \{\neg(\phi \mathrel{-\!\!3} (Ec \supset \psi c))\}$$

is inconsistent. In that case, there are $\gamma_1, \ldots, \gamma_m \in T_n$ such that

$$(\gamma_1 \wedge \ldots \wedge \gamma_m \wedge \neg(\phi \mathrel{-\!\!3} \forall x \psi x)) \supset (\phi \mathrel{-\!\!3} (Ec \supset \psi c))$$

But this sentence itself is a strict condition and c appears nowhere outside $Ec \supset \psi c$. So by MUG and predicate logic and the fact that all of our intensional operators are normal:

$$(\gamma_1 \wedge \ldots \wedge \gamma_m \wedge \neg(\phi \mathrel{-\!\!3} \forall x \psi x)) \supset (\phi \mathrel{-\!\!3} \forall x \psi x)$$

But then by propositional logic:

$$\neg(\gamma_1 \wedge \ldots \wedge \gamma_m)$$

So the original set is inconsistent, which is contrary to assumption. ■

Once we have lemma 3.10.2, we can show that every strongly consistent set has a coherent network using the same argument as before. We then need to prove a new version of the repair lemma. The new proof turns out to be the same as the old proof, with the exception that we need to appeal to lemma 3.10.3 in place of lemma 3.9.3. This is need to ensure that we can always add a world as an appropriate witness without further expanding the language.

Lemma 3.10.3. *If there is a superb set of sentences S and $\Diamond\phi \in S$, then there is a superb set of sentences extending $T_0 = \{\psi \mid \Box\psi \in S\} \cup \{\phi\}$ in the same language as S.*

Proof. Let S be a superb set of sentences. We then construct a superb set of sentences T extending T_0 by building a chain of sets just like the one in lemma 3.10.2. The only difference is that we set $T_0 = \{\psi \mid \Box\psi \in S\} \cup \{\phi\}$ and do not expand the language. The main concern, then, is that the first recursive clause is not well defined, since there may not be any such constant c. What we need to show is that this concern is misplaced.

Suppose for reductio that there is no such c. That is, suppose that $T_n \cup \{\neg(\phi \mathrel{\Box\!\!\rightarrow} \forall x\psi x)\}$ is consistent, but that no such T_{n+1} is consistent. This means that for every c, there are $\Box\gamma_1, ..., \Box\gamma_n \in S$ such that:

$$(\gamma_1 \wedge ... \wedge \gamma_n) \wedge \neg(\phi \mathrel{\Box\!\!\rightarrow} \forall x\psi x) \supset (\phi \mathrel{\Box\!\!\rightarrow} (Ec \supset \psi c))$$

It then follows by the normality of the box operator that

$$(\Box\gamma_1 \wedge ... \wedge \Box\gamma_n) \supset \Box(\neg(\phi \mathrel{\Box\!\!\rightarrow} \forall x\psi x) \supset (\phi \mathrel{\Box\!\!\rightarrow} (Ec \supset \psi c)))$$

for all constants c. But then since

$$\Box\gamma_1 \wedge ... \wedge \Box\gamma_n \in S$$

and S is closed under logical entailment, we have

$$\Box(\neg(\phi \mathrel{\Box\!\!\rightarrow} \forall x\psi x) \supset (\phi \mathrel{\Box\!\!\rightarrow} (Ec \supset \psi c))) \in S$$

for all c. Because S is strictly omega complete, this entails that:

$$\Box(\neg(\phi \mathrel{\Box\!\!\rightarrow} \forall x\psi x) \supset (\phi \mathrel{\Box\!\!\rightarrow} \forall x\psi x)) \in S$$

But then

$$\neg(\phi \mathrel{\Box\!\!\rightarrow} \forall x\psi x) \supset (\phi \mathrel{\Box\!\!\rightarrow} \forall x\psi x) \in T_n$$

which contradicts the original assumption that $T_n \cup \{\neg(\phi \mathrel{\Box\!\!\rightarrow} \forall x\psi x)\}$ is consistent. So there is always a constant c of the kind required. ∎

Once we have a new version of the repair lemma, it follows that every strongly consistent set has a perfect network by lemma 3.9.5, which we proved in the last section. But then we have the result that we wanted.

Theorem 3.10.1 (Completeness)**.** *Every strongly* $\mathbf{Q_2}$*-consistent set has a two-dimensional quantified Kripke model.*

Proof. Immediate from the preceding. ∎

3.11. Simplifying Matters

In the last section, we shoed that $\mathbf{Q_2}$ is complete using MUG and PD. What we are going to do now is show that these principles can be derived in $\mathbf{QTB_2}$, and so do not need to be taken as basic. This is the major step toward showing that $\mathbf{QTB_2}$ is complete with respect to the class of reflexive and symmetric frames. The rest of the argument is just a reprisal of the last two section.

Lemma 3.11.1. *The following rules are all valid in* $\mathbf{QTB_2}$.

Export $\phi \supset \Box\psi \Rightarrow \otimes\Diamond\otimes\phi \supset \psi$

$\phi \supset \otimes\psi \Rightarrow \otimes\phi \supset \psi$

$\phi \supset {\uparrow}\psi \Rightarrow \phi \supset \Box(\mathsf{Id} \supset \psi)$

Import $\otimes\Diamond\otimes\phi \supset \psi \Rightarrow \phi \supset \Box\psi$

$\otimes\phi \supset \psi \Rightarrow \phi \supset \otimes\psi$

$\phi \supset \Box(\mathsf{Id} \supset \psi) \Rightarrow \phi \supset {\uparrow}\psi$

Proof. Using TB, we can show the following, then use it to derive both the first export rule and the first import rule:

$$\otimes\Diamond\otimes\Box\phi \supset \phi$$

The second export rule and the second import rule are immediate by A2 and A4. The third import rule follows from T, which can be derived from TB. And finally, the third export rule can be derived using $\mathbf{K_2}$. The basic idea is to first show that

$$\Box\otimes\uparrow\phi \supset \Box\otimes(\mathsf{Id} \supset \phi)$$

using A7 and the fact that the relevant operators are normal. A5 and propositional logic then give us:

$$\phi \supset \Box\otimes(\mathsf{Id} \supset \phi)$$

At this point, we just have to eliminate the swap operator from the consequent. This can be done by showing that

$$\otimes(\mathsf{Id} \supset \phi) \equiv (\mathsf{Id} \supset \phi)$$

and using substitution. We then have

$$\uparrow\phi \supset \Box(\mathsf{Id} \supset \psi)$$

But in that case, we also have the validity of the third export rule. ■

Lemma 3.11.2. *MUG is valid in* $\mathbf{QTB_2}$.

Proof. Take any strict conditional. We can assume that it has no leading operators because, if it did, we could just use the logically equivalent $\top \supset (\phi \prec\!\!\!\!\!\!\; \psi)$. This means that we now have a sentence

$$\phi_1 \supset \ldots (\phi_2 \supset \ldots \psi \ldots)$$

with any number of operators in front of the first embedded conditional. For concreteness, we can think of this as:

$$\phi_1 \supset \Box\otimes\downarrow(\phi_2 \supset \ldots\ \psi\ \ldots\ )$$

Repeated applications of the export rules let us move the operators out of the consequent.

$$\Diamond\otimes\otimes\Diamond\otimes\phi_1 \supset (\mathsf{Id} \supset (\phi_2 \supset \ldots\ \psi\ \ldots\ ))$$

Propositional logic then gives us:

$$(\Diamond\otimes\otimes\Diamond\otimes\phi_1 \wedge \mathsf{Id} \wedge \phi_2) \supset\ \ldots\ (\ \ldots\ \psi\ \ldots\ )$$

Now at this point, there could be any of number of further embedded conditions with any number of operators in front of them. But if so, we can just repeat the process just described until there are none. We can assume, then, without loss of generality, that there are no further embedded conditionals. This gives us:

$$(\Diamond\otimes\otimes\Diamond\otimes\phi_1 \wedge \mathsf{Id} \wedge \phi_2) \supset \psi$$

But now what we have is just an ordinary material conditional. So we can universally generalize using predicate logic.

$$(\Diamond\otimes\otimes\Diamond\otimes\phi_1 \wedge \mathsf{Id} \wedge \phi_2) \supset \forall x\psi$$

Propositional logic and repeated application of the import rules then let us put all the operators back where we found them.

$$\phi_1 \supset \Box\otimes\downarrow(\phi_2 \supset \forall x\psi)$$

So MUG is valid in $\mathbf{QTB_2}$. ■

Lemma 3.11.3. *PI and $PD_{\otimes}$ are valid in* $\mathbf{FK_2}$*. Moreover, $PD_{\uparrow}$ entails $PD_{\Box}$ in* $\mathbf{FK_2}$ *and $PD_{\uparrow}$ is valid in* $\mathbf{QTB_2}$*.*

Proof. The proof that $PI_{\Box}$ is valid is the same as in the one-dimensional case. Since that proof only requires QPL and RN, the other two cases are the same. We can then derive $PD_{\otimes}$ from $PI_{\otimes}$ and A2 using propositional logic. This leaves us with only two remaining proofs.

To show: $t \neq s \supset \uparrow(t \neq s) \Rightarrow t \neq s \supset \Box(t \neq s)$

(1)	$(t \neq s) \supset \uparrow(t \neq s)$	
(2)	$\uparrow(t \neq s) \supset \Box\otimes\uparrow(t \neq s)$	A5
(3)	$(t \neq s) \equiv \uparrow(t \neq s)$	1, PI, A1
(4)	$(t \neq s) \equiv \otimes(t \neq s)$	$PD_{\otimes}$, PI, A2
(5)	$(t \neq s) \supset \Box(t \neq s)$	2, 3, 4, sub

To show: $t \neq s \supset \uparrow(t \neq s)$

(1)	$\uparrow(t = s) \supset \uparrow(t = s)$	PL
(2)	$\uparrow(t = s) \supset \uparrow\Box\otimes(t = s)$	1, PI
(3)	$\uparrow(t = s) \supset \Box\otimes(t = s)$	2, A6
(4)	$\uparrow(t = s) \supset (t = s)$	3, TB
(5)	$(t \neq s) \supset \uparrow(t \neq s)$	4, PL, A1

■

Theorem 3.11.1 (Completeness)**.** *Every* $\mathbf{QTB_2}$*-consistent set has a two-dimensional Kripke model.*

Proof. By lemma 3.11.2 and lemma 3.11.3 and an argument like the one from section 3.7 ■

Earlier, we noted that you might want to drop the Id operator. If so, then the derivation of MUG in lemma 3.11.2 will not go through, since the export rule for $\uparrow$ makes essential use of Id. Fortunately, when identity is dropped, there is another strategy available. This second strategy relies on accepting $\mathrm{BF}_{\uparrow}$. This axioms says, basically, that everything that exists at a world w from its own perspective also exists at w from the perspective of any other world v.

Lemma 3.11.4. *The following rules are all valid in* $\mathbf{QBT_2}$ *when the identity operator is dropped from the language.*

Export $\phi \supset \Box\psi \Rightarrow \otimes\Diamond\otimes\phi \supset \psi$

$\phi \supset \otimes\psi \Rightarrow \otimes\phi \supset \psi$

$\phi_1 \supset \uparrow(\phi_2 \supset \psi) \Rightarrow \phi_1 \supset (\uparrow\phi_2 \supset \uparrow\psi)$

$\phi \supset \uparrow\Box\phi \Rightarrow \phi \supset \Box\phi$

$\phi \supset \uparrow\otimes\phi \Rightarrow \phi \supset \uparrow\phi$

$\phi \supset \uparrow\uparrow\phi \Rightarrow \phi \supset \uparrow\phi$

Import $\otimes\Diamond\otimes\phi \supset \psi \Rightarrow \phi \supset \Box\psi$

$\otimes\phi \supset \psi \Rightarrow \phi \supset \otimes\psi$

$\phi_1 \supset (\uparrow\phi_2 \supset \uparrow\psi) \Rightarrow \phi_1 \supset \uparrow(\phi_2 \supset \psi)$

$\phi \supset \Box\phi \Rightarrow \phi \supset \uparrow\Box\phi$

$\phi \supset \uparrow\phi \Rightarrow \phi \supset \uparrow\otimes\phi$

$\phi \supset \uparrow\phi \Rightarrow \phi \supset \uparrow\uparrow\phi$

Proof. Immediate by lemma 3.11.1 and the basic rules and axioms. ■

Lemma 3.11.5. *MUG is valid in* $\mathbf{QTB_2}$ *when the identity operator is dropped and the axiom* $BF_\uparrow$ *is added.*

Proof. As before, take any strict conditional. We can assume that it has no leading operators, for the reason described earlier. We thus have a sentence of the form

$$\phi_1 \supset \ldots (\phi_2 \supset \ldots \psi \ldots)$$

with any number of operators in front of the first embedded conditional. For concreteness, we can think of this as:

$$\phi_1 \supset \uparrow\otimes\uparrow\Box\uparrow(\phi_2 \supset \ldots \psi \ldots)$$

Our new export rules let us eliminate all the arrows and swaps in front of the box.

$$\phi_1 \supset \Box\uparrow(\phi_2 \supset \ldots \psi \ldots)$$

The export rule for the box lets us move it out of the way as well.

$$\otimes\Diamond\otimes\phi_1 \supset \uparrow(\phi_2 \supset \ldots \psi \ldots)$$

Now we have an up operator directly in front of a material condition. Using the third export rule, we can distribute this operator. Propositional logic then gives us:

$$(\otimes\Diamond\otimes\phi_1 \wedge \uparrow\phi_2) \supset \uparrow \ldots (\ \ldots \psi \ldots)$$

At this point, the inner $\uparrow$ operator may be prefacing any number of operators. There may also be any number of further embedded conditionals. But in either case, we can just repeat the process just described to remove them. So we can assume, without loss of generality, that there are no further operators or embedded material conditionals. The result is:

$$(\otimes\Diamond\otimes\phi_1 \wedge \uparrow\phi_2) \supset \uparrow\psi$$

At this point, we can use predicate logic to universally generalize and move the universal quantified directly outside the innermost up operator.

$$(\otimes\Diamond\otimes\phi_1 \wedge \uparrow\phi_2) \supset \forall x\uparrow\psi$$

Using the axiom $BF_{\uparrow}$, we can then move that quantifier inside the up operator.

$$(\otimes\Diamond\otimes\phi_1 \wedge \uparrow\phi_2) \supset \uparrow\forall x\psi$$

At this point, we can then use the import rules to put all the operators back where we found them. The result is:

$$\phi_1 \supset \uparrow\otimes\uparrow\Box\uparrow(\phi_2 \supset \forall x\psi)$$

So MUG is valid in $\mathbf{QTB_2}$ when the identity operator is dropped from the language and the axiom $BF_{\uparrow}$ is added. ■

Theorem 3.11.2 (Completeness)**.** *Every* $\mathbf{QTB_2}$*-consistent set has a two-dimensional Kripke model when the axiom* $BF_{\uparrow}$ *is added and the* Id *operator is dropped.*

Proof. By lemma 3.11.3 and lemma 3.11.5 and an argument like the one from section 3.7 ■

Bibliography

Baker, David. 2015. “Some Consequences of Physics for the Comparative Metaphysics of Quantity.”

Belot, Gordon. 2011. *Geometric Possibility.* Oxford ; New York: Oxford University Press.

Berntson, Daniel. 2019. “Modal Ontological Relativism.”

Blackburn, Patrick, Maarten de Rijke, and Yde Venema. 2010. *Modal Logic.* 4. print. Cambridge Tracts in Theoretical Computer Science 53. Cambridge: Cambridge Univ. Press.

Boylan, David, and Ginger Schultheis. 2018. “Strengthening Principles and Counterfactual Semantics.”

Burgess, John P. 2009. *Philosophical Logic.* Princeton Foundations of Contemporary Philosophy. Princeton: Princeton University Press.

Burgess, John P., and Gideon A. Rosen. 1997. *A Subject with No Object: Strategies for Nominalistic Interpretation of Mathematics.* Oxford : New York: Clarendon Press ; Oxford University Press.

Chandler, Hugh S. 1976. “Plantinga and the Contingently Possible.” *Analysis* 36 (2): 106–109.

Cresswell, M. J. 1990. *Entities and Indices.* Dordrecht: Springer Netherlands.

Dasgupta, Shamik. 2014. “On the Plurality of Grounds.” 14 (20): 28.

Field, Hartry. 1980. *Science Without Numbers.* Princeton University Press.

———. 1984. “Can We Dispense with Space-Time?” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1984, no. 2 (January): 33–90.

Fine, Kit. 2005. *Modality and Tense: Philosophical Papers.* Oxford: Clarendon Press.

Fintel, Kai von. 2001. “Counterfactuals in a Dynamic Context.” In *Ken Hale: A Life in Language,* edited by Michael Kenstowicz, 123–152. MIT Press.

Garson, James W. 2001. “Quantification in Modal Logic.” In *Handbook of Philosophical Logic,* edited by D. M. Gabbay and F. Guenthner, 267–323. Dordrecht: Springer Netherlands.

Gillies, Anthony S. 2007. “Counterfactual Scorekeeping.” *Linguistics and Philosophy* 30 (3): 329–360.

Gupta, Haragauri Narayan. 1965. “Contributions to the Axiomatic Foundations of Geometry.” PhD Thesis, University of California, Berkeley.

Hazen, Allen. 1976. “Expressive Completeness in Modal Language.” *Journal of Philosophical Logic* 5 (1): 25–46.

Helle, Benj, Adam Murray, and Jessica Wilson. 2018. “Relativized Metaphysical Modality: Index and Context.”

Hughes, G. E., and M. J. Cresswell. 1996. *A New Introduction to Modal Logic.* London ; New York: Routledge.

Kemp, Gary. 2000. "The Interpretation of Crossworld Predication." *Philosophical Studies* 98 (3): 305–320.

Kratzer, Angelika. 1981. "Partition and Revision: The Semantics of Counterfactuals." *Journal of Philosophical Logic* 10 (2): 201–216.

Lange, Marc. 2009. *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature.* Oxford: Oxford Univ. Press.

Lewis, David. 1968. "Counterpart Theory and Quantified Modal Logic." *Journal of Philosophy* 65 (5): 113–126.

———. 1981. "Ordering Semantics and Premise Semantics for Counterfactuals." *Journal of Philosophical Logic* 10 (2): 217–234.

———. 2001. *On the Plurality of Worlds.* Malden, Mass: Blackwell Publishers.

Mackay, J. 2013. "Quantifying over Possibilities." *Philosophical Review* 122, no. 4 (January 1): 577–617.

Manders, Kenneth L. 1982. "On the Space-Time Ontology of Physical Theories." *Philosophy of Science* 49 (4): 575–590.

Marx, Maarten, and Yde Venema. 1997. *Multi-Dimensional Modal Logic.* Dordrecht: Springer Netherlands.

Milne, Peter. 1992. “Modal Metaphysics and Comparatives.” *Australasian Journal of Philosophy* 70, no. 3 (September): 248–262.

Morton, Adam. 1984. “Comparatives and Degrees.” *Analysis* 44, no. 1 (January 1): 16–20.

Mundy, Brent. 1986. “The Physical Content of Minkowski Geometry.” *The British Journal for the Philosophy of Science* 37 (1): 25–54.

———. 1987. “The Metaphysics of Quantity.” *Philosophical Studies* 51, no. 1 (January): 29–54.

Murray, Adam, and Jessica Wilson. 2012. “Relativized Metaphysical Modality*.” In *Oxford Studies in Metaphysics Volume 7,* edited by Karen Bennett and Dean W. Zimmerman, 189–226. Oxford University Press, October 25.

Nolan, Daniel Patrick. 2002. *Topics in the Philosophy of Possible Worlds.* Studies in Philosophy. New York: Routledge.

Parsons, Terence. 2014. *Articulating Medieval Logic.* Oxford: Oxford University Press.

Pieri, Mario. 1908. “La Geometria Elementare Istituita Sulle Nozioni Di Punto e Sfera.” *Matematica e di Fisica della Società Italiana delle Scienze* 15:345–450.

Pollock, John L. 1975. “Four Kinds of Conditionals.” *American Philosophical Quarterly* 12 (1): 51–59.

———. 1976a. *Subjunctive Reasoning.* Reidel.

———. 1976b. “The ?Possible Worlds? Analysis of Counterfactuals.” *Philosophical Studies* 29, no. 6 (June): 469–476.

Pollock, JohnL. 1981. “A Refined Theory of Counterfactuals.” *Journal of Philosophical Logic* 10, no. 2 (May).

Quine, W. V. 1961. “On What There Is.” In *From a Logical Point of View,* edited by Tim Crane and Katalin Farkas, 21–38. Harvard University Press.

Salmon, Nathan Ucuzoglu. 1979. “How Not to Derive Essentialism from the Theory of Reference.” *The Journal of Philosophy* 76, no. 12 (December): 703.

Stalnaker, Robert C. 1968. “A Theory of Conditionals.” In *Studies in Logical Theory (American Philosophical Quarterly Monographs 2),* edited by Nicholas Rescher, 98–112. Blackwell.

———. 1981. “A Defense of Conditional Excluded Middle.” In *Ifs,* edited by William Harper, Robert C. Stalnaker, and Glenn Pearce, 87–104. Reidel.

Stalnaker, Robert C., and Richmond H. Thomason. 1970. “A Semantic Analysis of Conditional Logic.” *Theoria* 36 (1): 23–42.

Tarski, Alfred. 1952. “A Decision Method for Elementary Algebra and Geometry.” *Journal of Symbolic Logic* 17 (3): 207–207.

———. 1959. “What Is Elementary Geometry?” In *Studies in Logic and the Foundations of Mathematics,* 27:16–29. Elsevier.

Van Fraassen, Bas C. 1974. “Hidden Variables in Conditional Logic.” *Theoria* 40 (3): 176.

Veblen, Oswald. 1904. “A System of Axioms for Geometry.” *Transactions of the American Mathematical Society* 5, no. 3 (March 1): 343–343.

Walters, Lee. 2009. “Morgenbesser’s Coin and Counterfactuals with True Components.” *Proceedings of the Aristotelian Society, New Series* 10:365–379.

Wehmeier, Kai Frederick. 2003. “World Travelling and Mood Swings.” In *Foundations of the Formal Sciences II,* edited by Benedikt Löwe, Thoralf Räsch, and Wolfgang Malzkorn. Kluwer.