

Defining Determinism

HP Halvorson, JB Manchak, JO Weatherall

March 7, 2025

Abstract

Determinism is the thesis that the past determines the future, but efforts to define it precisely have exposed deep methodological disagreements. Standard possible-worlds formulations of determinism presuppose an "agreement" relation between worlds, but this relation can be understood in multiple ways – none of which is particularly clear. We critically examine the proliferation of definitions of determinism in the recent literature, arguing that these definitions fail to deliver clear verdicts about actual scientific theories. We advocate a return to a formal approach, in the logical tradition of Carnap, that treats determinism as a property of scientific theories, rather than an elusive metaphysical doctrine.

We highlight two key distinctions: (1) the difference between qualitative and "full" determinism, as emphasized in recent discussions of physics and metaphysics, and (2) the distinction between weak and strong formal conditions on the uniqueness of world extensions. We argue that defining determinism in terms of metaphysical notions such as haecceities is unhelpful, whereas rigorous formal criteria – such as Belot's D1 and D3 – offer a tractable and scientifically relevant account. By clarifying what it means for a theory to be deterministic, we set the stage for a fruitful interaction between physics and metaphysics.

1 Determinism: Capturing the intuition

The thesis of determinism seems easy enough to state:

Determinism: The past determines the future.

Unfortunately, stated this way, the thesis is uninformative, because the term “determines” on the right is just a variant of the term we are trying to define. Can we do better than this?

In the world of post-1970s analytic metaphysics, it seems that all theses may be cashed out in terms of possible worlds. That lingo apparently provides a sharper formulation of the thesis:

Determinism: For any two possible worlds W, W' , if W and W' agree on the past, then W and W' agree on the future.

This possible-worlds definition gives a quasi-mathematical gloss to the word “determines”: there is a well-defined function from past segments to possible worlds. It seems like we have made good progress already in cashing out the notion of determinism. We have replaced the opaque word “determines” with what appears to be a statement about existing things and a clear relation (“agree”) between them.

But there is a problem. The relation “agree” has turned out to be a point of great contention among philosophers. To illustrate with a simple example:

(Q) Let W be a world with one particle on the left and an identical particle on the right. Let W' be the world in which the two particles are swapped. Do W and W' agree?

Indecision about Q can infect our judgment about whether determinism holds in concrete examples. For example, we can imagine that W and W' are exactly identical at all times leading up to now. If W and W' are judged to agree at the present time, then determinism holds; if W and W' are judged not to agree at the present time, then determinism fails. It seems that the fairly clear notion of determinism has become clouded by esoteric questions about what it means for possible worlds to agree.

One’s initial reaction might be that these kinds of examples only show that the notion of determinism was less clear than we imagined it was. But what are we to say about the fact that scientists seem confident in their judgments about whether physical theories are deterministic? For example, there is a genuine difference between deterministic and stochastic equations of motion. Or, perhaps more controversially, but still to the point: it seems quite clear that there is a difference vis-a-vis determinism between quantum mechanics and the theories that went before it. Scientists appear to be able make these distinctions reliably without having a theory of when possible worlds “agree”. What is behind those judgments?

Our basic claim in this paper is that whether or not a physical theory is deterministic can be seen by analyzing the mathematical structure of the theory. Talk of possible worlds is helpful only insofar as it sometimes functions as a substitute for talk of models of a theory; for most purposes, it introduces more problems than it solves. And we do not need a different metaphysical framework to replace Lewisian possible worlds, either. Mathematics is perfectly adequate for this task.

Our arguments runs against the grain of the past three decades of work on determinism in physics (and metaphysics), which has neared consensus that there can be no adequate “formal” definition of determinism. A theory, it is said, can only be deterministic or indeterministic under an “interpretation”. This is mistaken, at least as usually intended, both because it relies on confused ideas about “interpretation” and because a perfectly adequate formal criterion of determinism is already available—and has been for thirty years. We will argue that challenges for this criterion that have previously been discussed do not show that “interpretation” is needed; rather, they highlight the importance of precisely formulating a theory, as small differences in what we take a theory to be can lead to different judgments about whether that theory is deterministic.

2 The fate of formal definitions

The young Bertrand Russell frequently contrasted the clarity of nineteenth century mathematics with the opacity of Hegelian metaphysics. He pointed out that mathematicians had provided sharp definitions of the traditional number systems as well as general concepts such as infinite sets and continuous functions.

What is infinity? If any philosopher had been asked for a definition of infinity, he might have produced some unintelligible rigmarole, but he would certainly not have been able to give a definition that had any meaning at all. Twenty years ago, roughly speaking, Dedekind and Cantor asked this question, and, what is more remarkable, they answered it. They found, that is to say, a perfectly precise definition of an infinite number or an infinite collection of things. (Russell, 1901, p 92)

Russell then proposed that philosophers should model their method on that

of the nineteenth century mathematicians. Obviously, Russell's vision was a shaping force in analytic philosophy, beginning with Carnap's attempts to "explicate" the concepts of the natural sciences (see Leitgeb and Carus, 2020).

In *Logical Syntax of Language*, Carnap suggests that some deep and murky metaphysical theses correspond to precise statements about the structure of scientific theories. In particular, with regard to determinism, he says:

The opposition between the determinism of classical physics and the probability determination of quantum physics concerns a syntactical difference in the system of natural laws, that is, of the P-rules of the physical language. (Carnap, 1937, p 307)

More concretely, Carnap suggests that the metaphysical doctrine:

Every process is univocally determined by its causes

corresponds to a formal property of a scientific theory:

For every particular physical sentence φ , there is for any time coordinate t , which has a smaller value than the time coordinate which occurs in φ , a class Γ of particular sentences with t as time coordinate, such that φ is a P-consequence of Γ .

The formal property described here may seem more opaque than the original metaphysical doctrine. Nonetheless, unlike the original metaphysical doctrine, whether a theory is deterministic in the latter sense is something that could be checked by a mathematician, so long as the relevant theory has been specified in a mathematically clear fashion.

Even after Carnap's methods fell under attack by Quine, some notable philosophers continued to think of determinism as a formal property of scientific theories. J.J.C. Smart claims that:

A perfectly precise meaning can be given to saying that certain theories are deterministic or indeterministic (for example that Newtonian mechanics is deterministic, quantum mechanics indeterministic), but our talk about actual events in the world as being determined or otherwise may be little more than a reflection of our faith in prevailing types of physical theory. (Smart, 1961, p 294)

But change was in the wind, and by the 1970s, one begins to find a different kind of definition of determinism. The most influential of these “metaphysical” definitions of determinism is from David Lewis:¹

A system of laws of nature is Deterministic iff no two divergent worlds both conform perfectly to the laws of that system. Second, a world is Deterministic iff its laws comprise a Deterministic system. Third, Determinism is the thesis that our world is Deterministic. (Lewis, 1983, p 360)

Lewis’ definition was put front and center in philosophy of physics by Earman (1986), and it has ever-since served as the backdrop for the “hole argument” in General Relativity (see Butterfield, 1989; Pooley, 2021). More generally, Lewis’ diverging-worlds conception of determinism informs a wide range of discussions in analytic philosophy.

What was originally an inclination to do philosophy in a different way (than the logical positivists) soon developed into theoretical arguments against formal approaches. For example, Belot (1995a) argues that determinism cannot possibly be defined in anything like the way that Carnap and Smart proposed, because it is not a *formal* property of (mathematical) theories:

The first point that I would like to make is that determinism cannot be a *formal* property of physical theories. (Belot, 1995a, p 88, emphasis in original)

Belot’s position here is part of a trend among analytic metaphysicians and metaphysically-oriented philosophers of science away from formal definitions. For example, regarding formal definitions of equivalence of theories, Sider claims that, “the purely formal approach is a nonstarter” (2020, p 180), and, “purely formal accounts fail because they entirely neglect meaning” (2020, p 181). A similar complaint against formal definitions of theoretical equivalence is voiced by Kevin Coffey:

The challenges posed by the two puzzles are not unique to formal approaches, but I think we should be particularly pessimistic about the prospects of formal approaches meeting those challenges. (Coffey, 2014, p 834)

¹Lewis credits Montague (1962) with a similar definition, but notes that Montague does not focus on metaphysical issues. We won’t discuss Montague’s paper in detail, but we suggest that his argument against a “syntactic” definition of determinism is unconvincing, and is based on a false dilemma between syntactic and semantic definitions.

And Trevor Teitel argues that formal definitions are of little interest for philosophical investigations:

I will investigate various views one might hold about the non-mathematical significance of these formal criteria, and argue that none is tenable. My tentative conclusion is that formal criteria are of limited non-mathematical interest. (Teitel, 2021, p 4120)

What justifies this widespread rejection of formal methods? Sider, Coffey, and Teitel make different arguments from Belot, and from one another. But they also put great weight on the importance of *interpretation* or, as Sider says, *meaning*. We will focus on Belot's argument against formal definitions of determinism. When we see why that argument fails, it will be clear why we would reject other arguments for the same kind of view. And then, with formal methods rehabilitated, we will turn to a formal analysis of other arguments in the literature on determinism.

3 Interpretation is a formal matter

Belot's argument for the claim that determinism cannot be a formal property of theories involves an example. He presents a set of equations – Maxwell's equations, describing classical electromagnetism, written in a particular way – that he claims are standardly understood to be part of a deterministic theory, but which may yet also be part of an indeterministic one. The difference, he says, comes down to interpretation.

This completes the argument: determinism cannot be a formal property of theories, because the same theory may be deterministic or indeterministic, depending on how it is interpreted. (Belot, 1995a, p. 88)

The argument as stated has a suppressed premise: that interpretation is not a formal matter. But we deny this premise, twice over. First, the claim that interpretation is not a formal matter takes interpretation to involve a re-negotiation of the relationship between language and reality. Second, and relatedly, the claim that interpretation is not a formal matter fails to see that interpretation of a theory very often calls for a formal precisification of that theory.

We take the second claim first, because Belot’s own argument illustrates the point. Belot’s example concerns two theories. One of these theories concerns the properties of particles called “blips”, which are represented by two scalar fields; the other concerns the properties of electromagnetic fields, whose behavior can be described, in a somewhat more complicated way, by two scalar fields that happen to satisfy the “same” (syntactic) equations. Belot argues the first of these theories is indeterministic, whereas the second is deterministic. This is so even though they involve the same equations.

This argument is surprising on its face, since no one, we claim, would argue that one can infer whether a theory is deterministic by looking at just fragments of the theory—nor would we say that theories are individuated by equations that appear among their axioms. Instead, we must look at the theories as a whole. Belot does not fully articulate these theories in a formal way. But he describes them with sufficient clarity that one can see why they are mathematically distinct—and it is precisely those mathematical differences that support his judgments about whether they are deterministic.

Of course, Belot would dispute that these differences between these theories are *merely* formal (or mathematical), because he maintains that interpretation plays some role. What does he mean by “interpretation”? He poses and answers the question directly.

What *is* an interpretation of a theory? An interpretation of a theory is a story that you tell about the theory. . . . An interpretation is a correspondence between bits of models of the theory and bits of physical situations: between initial value constraints, variables and differential equations on the one hand, and instantaneous states, physical entities, properties, relations, etc., and laws of nature on the other. (Belot, 1995a, p 92)

This idea is widespread in late twentieth century analytic philosophy. But it should give us pause, because it involves an odd mix of two very different things. The first is the *formal* theory of interpretation, developed by Tarski et al. six decades before, which explicates an interpretation of a formal language as an assignment of set-theoretic extensions to meaningless symbols. Belot’s talk of the “models” of a theory invokes this tradition.

But Belot also motions towards a much woolier, and apparently older, sense of “interpretation” when he speaks of a correspondence with the world. In doing so, Belot has re-doubled the notion of interpretation. This brings us back to our first reason to reject Belot’s suppressed premise. In its Tarskian

sense, an interpretation of a theory *is* a model of that theory. But now Belot speaks of interpretation as, “a correspondence between bits of models of a theory and bits of physical situations”. In this case, the interpretations (models) of a theory are being interpreted. But what does that mean? How would we produce or exhibit a “correspondence between bits of models of the theory and bits of physical situations”?

The answer to this question would require a theory of word/world relations that would take us far beyond the scope of this paper. For our purposes, it suffices to observe that Belot does not give us a solution to this problem either. Indeed, when it comes time for him to describe two theories that differ (only?) in interpretation, what we find is not a “correspondence between bits of models of the theory and bits of physical situations”, but rather some *words*, further elaborating some formal aspects of two theories he wishes us to understand. The exercise is all carried out on paper. What he does is layer “interpretation” on “interpretation”, by sketching relationships between symbols of one sort and symbols of another. And on reflection, given that he has written a paper, it is hard to see what his sense of interpretation could amount to aside from creating yet other models, and attaching bits of the old models to bits of the new models. So in fact, though he motions at something different, what he exhibits as interpretation does not give us some new way of crossing the word-world barrier; it merely connects things that lie on the word side, the formal side, to other things on that side.

One might well object here: how does *any* of our linguistic or symbolic practice have meaning, if all of our “interpretation” consists of layering models on models and words upon words? Surely, one might say, we do sometimes interpret in the woolier, and perhaps deeper, sense—and at least some of our words have meaning. Of course we accept this. But we suggest this line of thought just leads to a dual perspective on our position, entirely consistent with what we have said thus far. From this perspective, *all* of the formal, mathematical, linguistic, and symbolic structures that we employ, in ordinary life, in philosophy, and in science, have meaning, at least to some extent, because of the way they are embedded in our broader cultural and cognitive processes.

Dewar (2023) puts the point nicely.

We don’t begin our analysis of scientific theories by taking some mysterious equations carved on stone tablets and puzzling out what they might mean: theories are born as bearing all kinds of

semantic or interpretational relationships to our broader representational practice. ... [T]he problem is not how to comprehend an alien practice, but how to fully understand a practice which we already – at least to some extent – inhabit. (18)

Again, we will not attempt to give a story of how meaning works. Our point is that there is no line to draw between “merely” or “purely” formal methods or structures, between truly “meaningless symbols” or marks on a page, and any other symbols we might try to employ to make our ideas clear or communicate them to others. The idea from the formal theory of interpretation that we are giving meaningless symbols “meaning” by mapping them to set-theoretic structures is an idealization—as should be obvious, since *both* sides of the interpretation map consist of rich but thoroughly formal theories. Likewise, interpretation in the wild consists of using meaningful symbols to guide reasoning about other meaningful symbols. It is all just more of the same.

This point is illustrated by Belot’s own argument. Why does he think that this talk about elements of reality has any bearing on the question of whether these interpretations are deterministic? The crucial step occurs when, in describing the theories, he explains when putatively distinct models actually correspond to the same physical situations (described in different terms). This amounts to a disambiguation between two possible theories, given by specifying when models are isomorphic. According to one disambiguation (in fact, the standard one), electromagnetism has arrows between models with different vector potentials, and this theory is deterministic. According to another disambiguation, electromagnetism – or rather, the blip theory – does *not* have arrows between these models, and this theory is indeterministic. But everything here is formal: the interpretations of the theory, as well as the definition of determinism. Belot has not escaped from the circle of formalism and into a realm of inarticulable concreta; he has merely motioned toward the power of formal tools to capture conceptual distinctions. And we thank him for this: he shows clearly that what appears to be a single theory is in fact two, depending on how one formalizes it. We can call this explication of formal structure “interpretation” if we want – as Tarski did – but in that case, interpretation is a formal matter.

Our view of interpretation may be more modest than Belot’s, but it also shows why interpretation might involve *formal* work. The salient case here is the formal work of specifying the morphisms between the models — because

different accounts of the morphisms give different theories, one of which might be deterministic, while the other is indeterministic. We will presently see a more subtle example of where careful interpretation of a different kind, though still formal, can disambiguate between theories whose differences are difficult to detect through more conversational descriptions.

4 From explication to metaphysical speculation

With that defense – and explanation – of formal methods in mind, we now return to analyses of determinism downstream of Lewis. There are two directions in which Lewis’ (1983) definition of determinism has been developed. One is the quasi-mathematical approach introduced by Earman but criticized by Belot. We return to that below. First, we will consider the other direction of development. On this branch, Lewis’ definition has been suited out with various metaphysical distinctions — and especially an elusive distinction between “qualitative” and “full” agreement of possible worlds. These metaphysical adumbrations on Lewis’ definition have given rise to a significant literature which might seem, at first, to evince a rich interaction between physics and metaphysics. But as we will presently argue – bringing formal methods to bear – these appearances are misleading.

Consider Hawthorne’s (2006) proposed distinction between two senses in which the world might be deterministic.²

Qualitative Determinism: For all times t , there is no possible world which matches this world in its qualitative description up to t , and which has the same laws of nature as this world, but which doesn’t match this world in its total qualitative description. (239)

De Re Determinism: For all times t , there is no possible world which matches this world in its de re description up to t , and which has the same laws of nature as this world, but which doesn’t match this world in its total de re description. (239)

Hawthorne’s distinction seems to be nearly ubiquitous in the literature. For example, Teitel (2019) distinguishes laws that are qualitatively deterministic from laws that are fully deterministic. Similarly, Pooley (2021) claims

²A similar distinction can be found in (Melia, 1999), though not quite as sharply put.

that what he calls “substantialist” General Relativity is deterministic to a lower degree “Det2”, but not in the most eminent sense “Det1”.³ Dewar (2016; 2024) proposes a more subtle distinction between de dicto and de re determinism, which nonetheless aligns closely with these other distinctions.

	grade 1	grade 2
Hawthorne	qualitative	de re
Teitel	qualitative	full
Pooley	Det2	Det1
Dewar	de dicto	de re

Each of these authors identifies two grades of determinism — a lower one with a qualitative flavor, and a higher one with what might be called a “haecceitistic” flavor. The going assumption seems to be that the lower grade of determinism is good enough for the everyday business of science; but for the purposes of true metaphysical understanding, the higher grade is needed.

Clearly, the terminology here has become variegated. For simplicity, we will use Teitel’s “full determinism” to gather together the different senses of higher grade determinism.⁴ We will eventually argue that full determinism simply does not apply to most scientific theories, since they do not have names for all objects. We will also show that when full determinism does apply, it equivalent to our preferred (formal) definition of determinism. Appearances to the contrary arise due to confusion about (formal) interpretation and imprecision in the individuation of theories.

For possible worlds W, W' , let us write $W \sim_q W'$ if W and W' match in qualitative description, in Hawthorne’s sense (i.e. are “qualitatively identical” in Pooley’s sense). Let’s write $W \sim_{dr} W'$ if W and W' match in de re description, in Hawthorne’s sense (i.e. are “intrinsically identical” in Pooley’s sense). While we agree that it is possible to distinguish between grades of determinism, we will show that such distinctions cannot be based on some distinction between qualitative and de re equivalence of worlds. In the remainder of this section, we will try to give the distinction between \sim_q and

³We see no value to the qualifier “substantialist” here, since what Pooley has in mind is apparently just textbook General Relativity. We will return to this point below; for now, we will drop the qualifier.

⁴The term *de re deterministic* has been given a precise meaning in (Manchak, Barrett, et al., 2025) that is essentially equivalent to Belot’s D3, which we discuss below. As will become clear, understood this way de re determinism is not really about haecceitism at all.

\sim_{dr} a run for its money — concluding that it only gives intelligible answers for theories with names for all objects. In Section 5, we propose a distinction between grades of determinism that does not rely on a distinction between qualitative and de re equivalence of worlds.

4.1 Descriptions: Qualitative versus De Re

Hawthorne and Teitel explain the distinction between \sim_{q} and \sim_{dr} in terms of a distinction between types of propositions. Roughly speaking, $M \sim_{\text{q}} M'$ means that $M \models \varphi$ iff $M' \models \varphi$, for all qualitative propositions φ . And $M \sim_{\text{dr}} M'$ means that $M \models \varphi$ iff $M' \models \varphi$, for all de re propositions φ . Can we make rigorous sense of a distinction between these two types of propositions?

Hawthorne explains the distinction between two kinds of descriptions as follows:

The first — the *qualitative description* — says everything that can be said about the intrinsic character of that history with one exception: it cannot name individuals or otherwise encode haecceitistic information about which particular individuals are caught up in that segment of world history. The second — the *de re description* — includes the qualitative description and, in addition, all haecceitistic, singular information. (Hawthorne, 2006, p 239)

Granted, such a distinction works fine in everyday life: in some contexts where we have a name for an individual, we can give a specific description of what properties that individual has, or we can give a general description of something that has those properties. This is a distinction between levels of generality.

And yet, for well known reasons, the distinction is not absolute. For instance, sometimes qualitative descriptions can contain singular information. As Russell himself taught us, the sentence

(D) The present king of France is bald.

does not contain the name of an individual, but is, in some sense, about a particular individual. (Or should we say that it *could be* about a particular individual?) Of course, D also implies less specific propositions, such as “There is a bald thing.” Conversely, even names may not include all singular

information, at least in ordinary language:⁵ “James Weatherall” may refer to a philosopher of physics, or to his father, or his son, or to a former All-American football player, or to a retired British Vice-Admiral. In each of those cases, further descriptive information is needed to disambiguate the reference of the name.

Perhaps more importantly, whether some description contains singular information is apparently not a fact about the description itself, but rather about the thing or things described. For example, “Some nobleman is bald” could be about no individuals, or about one, or about many. And in any of these cases, it implies the less specific proposition “Somebody is bald”. On these grounds, we are skeptical about the idea that there is some significant distinction between descriptions that encode information about individuals, and descriptions that do not.

But let’s try harder. Recall that a description is supposed to be qualitative just in case it does not name any particular individual. We will now consider several proposals for how to make this idea precise. We assume that propositions are expressed, up to logical equivalence, by sentences. Thus, we assume that if sentences are logically equivalent, then either they both name an individual, or neither of them names an individual. If that were not the case, then “ x names an individual” would not be a property of the underlying proposition, but only of some specific syntactic representation of that proposition, and that would not help in defining a notion of qualitative sameness of worlds.

Proposal 1. A sentence φ names an individual just in case φ contains a name.

We have already seen arguments against proposals of this sort, but stated in these terms it fails for even simpler reasons, since it depends on a superficial features of a sentence that are not invariant under logical equivalence. Indeed, φ is logically equivalent to $\varphi \wedge (c = c)$, so by this proposal, every sentence names an individual.

Proposal 2. A sentence φ names an individual just in case there is a name c such that for any sentence φ' , if φ' is logically equivalent to φ , then φ' contains c .

⁵In first-order logic, we regiment name usage to require names to refer uniquely. But that is a choice in how we set up our semantics, and it does not apply in ordinary language.

This proposal would make it essentially impossible for sentences to name individuals. Indeed, one could define a predicate $\theta(x) \leftrightarrow (x = c)$, and then replace the de re sentence $p(c)$ with the definitionally equivalent $\exists!x(\theta(x) \wedge p(x))$.

Proposal 3. Given a background theory T , a predicate $\theta(x)$ names some individual relative to T just in case $T \vdash \exists!x\theta(x)$.

This condition is not strong enough. Even when $T \vdash \exists!x\theta(x)$, there can be a model M of T such that $M \models \theta(a)$, and another model M' of T such that $M' \models \neg\theta(a)$. So, $T \vdash \exists!x\theta(x)$ does not capture the idea that $\theta(x)$ names an individual.

We have tried, without any success, to give a formal (i.e. mathematical) account of the distinction between qualitative and de re propositions. Perhaps our failure is unsurprising, since the distinction between qualitative and de re descriptions was intended to be a distinction in how those descriptions relate to the world, i.e., it is meant to be a distinction in what those descriptions are *about*, whereas we have been talking only about mathematical objects (e.g. sentences of formal languages, models).

Here is another strategy. Perhaps one could simply assume that there is a relation $N(\varphi, x)$ of “naming” that can hold for propositions and concrete objects.

Proposal 4. A proposition φ is *qualitative* just in case $\neg\exists xN(\varphi, x)$.

Such a proposal has all the advantages of theft over honest toil. But to be more serious, we are still hung up on the practical concern of whether the distinction would actually help us decide whether specific scientific theories are deterministic, and saying that there is such a relation N does not tell us which sentences are qualitative and which are de re. Nor would this proposal provide any concrete guidance on the question of whether models are de re or qualitatively equivalent. Metaphysicians would be within their rights to say “there is such a relation N ”, but until they tell us how to detect that $\exists xN(\varphi, x)$ for a proposition φ , their proposal is of no interest for understanding the metaphysical implications of science.

Perhaps, you might think, we are straying too far from common sense. Surely the following claim, which one might find in a science textbook, names a concrete physical object:

(E) The perihelion of Mercury advances by approximately 43 arc-seconds per century more than what is predicted by Newtonian mechanics.

But for reasons related to our arguments in section 3, we think this rests on a linguistic confusion. Yes, the sentence involves a name in the colloquial sense. But the sentence most certainly does not involve “all singular, haecceitistic” information about the planet closest to the sun. It manages to convey the intended meaning only because of its integration with linguistic and scientific practice, our history and culture, established conventions, and so on. Put another way, the term “Mercury” does not do anything, it is people who do things. Names and sentences and propositions do not name things, it is people who name them. We may wish to include an abstract intermediary, such as a proposition, that a person uses to refer to something—and we might single out certain non-logical vocabulary as especially well-suited for playing a referential role in our languages. But without the person in the middle, the proposition does not stand in an intrinsic relation to any concrete thing.

We find that we have entered a wild and unfamiliar metaphysical territory, far outside of our comfort zone. We have no carefully worked out view of the metaphysics of propositions, and we do not intend to take a stance on these issues. More power to them who believe that they have a theory of propositions that will shed further light on the elusive distinction between qualitative and full equivalence of possible worlds.

What we do intend to take a stance on is that some definitions are more useful than others. For example, what does it mean to say that two sentences are synonymous? As Quine pointed out, it is perfectly correct, and perfectly useless, to say that two sentences are synonymous if they express the same proposition. The problem is that nobody has ever figured out that two sentences are synonymous by comparing each of them with a proposition. In the same way, nobody will ever determine if a proposition φ is de re by checking $\exists xN(\varphi, x)$; and nobody will ever figure out that a scientific theory is deterministic by checking possible worlds for qualitative and de re equivalence.

4.2 Equivalence relations on models

The attempt to distinguish qualitative from de re descriptions via a distinction in kinds of sentences did not go so well. Let’s approach the problems

from a different direction. Hawthorne gives us a motivating example to help distinguish the two senses of determinism.

Consider a symmetrical world where there is a pair of qualitatively identical ships, one in each symmetrical half. Suppose the laws dictated that exactly one of the ships would sink, but left it undetermined which. Qualitative determinism might still hold of such a world, since the qualitative description of a world in which one ship sank need not depart in any way from the qualitative description of a world in which the other did. (Hawthorne, 2006, p 243)

This is meant to be a case where there are two models M and M' that are qualitatively equivalent ($M \sim_q M'$) but not de re equivalent ($M \not\sim_{dr} M'$).

Here is an attempt to capture the example more precisely. Let Σ be a signature with a single predicate symbol $p(x)$ for “ x sinks”. Let T be the Σ -theory that says that there are exactly two things, and that exactly one of these things has property p . Then a model of this theory consists of a set with two elements, one of which has property p . Here we can write a model as a pair $\langle \{a, b\}, a \rangle$, where the first element is the domain of the model, and the second element is the singleton extension of p . Surely then we know the difference between qualitative and de re equivalence of models of T . For example, it seems obvious that $M = \langle \{a, b\}, a \rangle$ and $M' = \langle \{a, b\}, b \rangle$ are qualitatively equivalent (since something is p), but not de re equivalent (since different things are p in the two cases).

But why *not* think that the models M and M' are de re equivalent? Consider a possible world with a pair of objects, one of which is p . We can name the thing that is p in various ways: we could name it a , or we could name it b . So why not take $\langle \{a, b\}, a \rangle$ and $\langle \{b, a\}, b \rangle$ to be two different names for the exact same world? Nor will it help here to say: “but a and b are not names of things, they are the things themselves”. That is just a bit of nonsense, for reasons already discussed.

The thought that M and M' are not de re equivalent might be based on the following:

Proposal 5. We say that $M \sim_q M'$ just in case there is an isomorphism $f : M \rightarrow M'$. We say that $M \sim_{dr} M'$ just in case $M = M'$ in the sense of ZF set theory.

We ourselves are inclined to think that set-theoretic identity of models is of no significance, here; we think that isomorphism of models is the only significant standard of sameness.⁶ But we do not need to convince you of that stronger position. It is enough to point out that while isomorphism of models certainly captures a sense in which those models “agree on all qualitative propositions”, this is merely a consequence of the fact that isomorphic models agree on all propositions, qualitative or not. The relevant fact is:

Let $\varphi(x)$ be a Σ -formula. If $f : M \rightarrow N$ is a Σ -isomorphism, then for all $a \in |M|$, $M \models \varphi(a)$ iff $N \models \varphi(f(a))$,

which has the immediate consequence that

Let φ be a Σ -sentence. If $f : M \rightarrow N$ is a Σ -isomorphism, then $M \models \varphi$ iff $N \models \varphi$.

In other words, isomorphic Σ -structures agree on all Σ -sentences. If someone wishes to respond that the Σ -sentences are all qualitative propositions, then they need to give an account of what the de re propositions are supposed to be. How should this go? Proposal 5 makes it seem that *any* set-theoretic statement about M is a de re proposition. But that seems wildly implausible, because these set-theoretic statements include things like:

There is an $a \in |M|$, and a chain of length 10^{123} of inclusions:
 $\emptyset \in \dots \in a$.

Something seems very wrong with the idea of distinguishing models by their deep set-theoretic structure. The internal structure of elements of a model is irrelevant to the way that model is used to represent the world. So, we conclude, de re equivalence $M \sim_{\text{dr}} M'$ cannot be set-theoretic identity $M = M'$.

Perhaps what we need, then, is to distinguish two languages: a qualitative sublanguage Σ_{q} and a full language Σ_{dr} that has sufficiently many constant symbols – that is, names. For example, for describing the twin ships, we can take Σ_{q} to consist of the predicate symbol p , while Σ_{dr} also contains a name

⁶Reasons for thinking this in a related context are given by Weatherall, 2018 and Bradley and Weatherall, 2022; other reasons, related to the arguments in section 3, are discussed below. Ultimately, the issue is that insisting on set-theoretic identity of models involves layering interpretation on interpretation in a way that fundamentally confuses what the theory expresses.

c for one of the two ships. (In this case, we need only one name c , since the other ship is then named by $\lambda x(x \neq c)$ – that is, the set of things not named c .) For the general case, we will assume that Σ_{dr} contains as many names as there are real numbers, and we will implicitly restrict to consideration of models of cardinality no greater than the real numbers.

Proposal 6. Let T be a Σ_{dr} -theory. For models M, M' of T , we write $M \sim_{\text{dr}} M'$ just in case there is a Σ_{dr} -isomorphism $f : M \rightarrow M'$. We write $M \sim_{\text{q}} M'$ just in case there is a Σ_{q} -isomorphism $g : M \rightarrow M'$.

This proposal gets the right answer for the example of a pair of identical objects, where exactly one of the two has property $p(x)$, e.g. x is sinking at $t = 1$. The following two models of T are not Σ_{dr} -isomorphic, but their reductions to Σ_{q} are isomorphic.

	M	M'
p	a	b
c	a	a

We take it then that Proposal 6 provides a good definition of de re equivalence for models of theories with sufficiently many names. In this case, we can use de re equivalence of models to define full determinism:⁷

Full Determinism (FD): Let T be Σ_{dr} -theory. Then T is deterministic just in case for any models M, M' of T , and for any initial segments U of M and U' of M' , if $U \sim_{\text{dr}} U'$ then $M \sim_{\text{dr}} M'$.

This definition gives the right answers for the simple kinds of examples we have considered. For example, for the theory T that says that one of the two ships sinks at $t = 1$, the model M with $M \models p(c)$ is de re inequivalent to the model M' with $M' \models \neg p(c)$, even though their initial segments (which do not include the property p) are de re equivalent.

The question then is whether FD can be generalized to render judgments for real scientific theories, most of which do not provide names for elements in their models. In the remainder of this section, we will argue that FD does *not* generalize. But first we will clarify the sense in which a typical physical theory, such as General Relativity, lacks names for the objects it quantifies over. The key fact is the following:

⁷This definition of Full Determinism is just Belot's D1 for a theory in a signature Σ_{dr} with sufficiently many names. In other words, it is Qualitative Determinism for the name-enriched theory.

Fact. *Let M be a model of T . If every element of $|M|$ is the referent of some name, then the identity $1_M : M \rightarrow M$ is the only automorphism of M .*

This fact is a trivial consequence of the definition of a homomorphism of Σ -structures (see Hodges, 1993). That definition requires that $f : M \rightarrow M$ is a homomorphism only if $f(c^M) = c^M$, for each constant symbol $c \in \Sigma$. Hence, if for each $a \in |M|$ there is some $c \in \Sigma$ such that $c^M = a$, then $f(a) = f(c^M) = c^M = a$. Since this is true for all $a \in |M|$, $f = 1_M$.

Imagine now that GR has been formulated syntactically so that a relativistic spacetime M is a model of GR. Then there are models M that have non-trivial automorphisms; e.g., Minkowski spacetime has non-trivial automorphisms. It follows then from Fact that GR does not have names for all the elements of its models. Therefore, Proposal 6 does not define relations \sim_{dr} and \sim_{q} on relativistic spacetimes, and FD does not apply to GR.⁸

To generalize FD to theories without names for all objects, we would presumably need to generalize Proposal 6 to define relations \sim_{dr} and \sim_{q} on models. But we claim this cannot be done. Suppose then that T is a theory in a qualitative signature Σ_{q} . We assume that $M \sim_{\text{q}} M'$ whenever there is a Σ_{q} -isomorphism $f : M \rightarrow M'$. The question is whether there is some finer-grained equivalence relation \sim_{dr} on the models of T . We will presently show that there is none that meets a basic desideratum.

Consider again our running example: Σ_{q} has a single predicate symbol p , and T says that there are exactly two things, one of which is p . For domain $\{a, b\}$, there is a model M such that $M \models p(a)$, and a model M' such that $M' \models p(b)$. There is a Σ_{q} -isomorphism $f : M \rightarrow M'$, hence $M \sim_{\text{q}} M'$. We could then declare that M and M' are de re inequivalent. But can we continue this definition of \sim_{dr} to all other models of T ?

Consider another domain $\{\alpha, \beta\}$, and models N, N' with $N \models p(\alpha)$ and $N' \models p(\beta)$. The convention just adopted tells us that N and N' should be de re inequivalent. But what about M and N ? It seems that we are free to choose a convention. So set $M \sim_{\text{dr}} N$ and $M' \sim_{\text{dr}} N'$.⁹ The astute reader will see that this process can be continued, establishing two de re equivalence

⁸One could consider enriching GR to include names, but we postpone a discussion of that point to section 8.

⁹This convention is tantamount to choosing a fixed isomorphism between the domain of M and the domain of N , or what's equivalent, to fixing the extension of a constant symbol c in each model of T . The problem we will encounter below is that it's impossible to satisfy the constraint that if $c^M \in |N|$, then $c^M = c^N$.

classes of models of T . So far so good. But on reflection, we suspect that this route will be unattractive to any advocate for FD. The reason that these two equivalence classes will not be *haecceitistically consistent*.

Definition. Let M and N be models of a theory T . We say that M and N are *haecceitistically consistent* just in case for any $a \in |M| \cap |N|$, and for any formula $\varphi(x)$, $M \models \varphi(a)$ iff $N \models \varphi(a)$.

In other words, haecceitistically inconsistent models share an element a in common, but disagree on whether a has some property $\varphi(x)$. This kind of haecceitistic difference is precisely what is thought to happen in the hole argument, i.e. the models M and N disagree on the properties assigned to some spacetime point $a \in |M| \cap |N|$.

Proposition 1. Let \sim_{dr} be any equivalence relation on models of T such that:

1. Each model M is de re inequivalent with the model M' that has the same domain, but flips the extension of p .
2. For any models M, N , either $M \sim_{\text{dr}} N$ or $M \sim_{\text{dr}} N'$.

Then there are models M, N of T such that $M \sim_{\text{dr}} N$, but M and N are haecceitistically inconsistent.

Proof. Let $M = \{a, b\}$ with $M \models p(a)$, and let $M_0 = \{a, *\}$ with $M_0 \models \neg p(a)$. If $M \sim_{\text{dr}} M_0$, then we are finished, so assume that $M \sim_{\text{dr}} M'_0$. Let $M_1 = \{*, b\}$ with $M_1 \models p(b)$. If $M \sim_{\text{dr}} M_1$, then we are finished, so assume that $M \sim_{\text{dr}} M'_1$. Then $M'_0 \sim_{\text{dr}} M'_1$, but $M'_0 \models \neg p(*)$ and $M'_1 \models p(*)$. That is, M'_0 and M'_1 are de re equivalent, but haecceitistically inconsistent. \square

This result shows that that there is no (haecceitistically consistent) notion of de re equivalence of models that is finer-grained than isomorphism. Hence, Full Determinism is otiose as a criterion for judging whether actual scientific theories are deterministic. We could, of course, try some other way to extend Full Determinism to theories without names for all objects. For example, given a theory T , we could supplement the language with names, and then ask if the name-enriched theory T^+ satisfies FD. But no matter what T itself is like, we should not expect T^+ to be deterministic, since we have added many new properties without dynamical laws governing their evolution. (We return to this point in section 8.)

5 Lewis formalized

Distinguishing senses of determinism by distinguishing types of descriptions or equivalence classes of models did not get us very far. But in fact, we think something very much in the spirit of the distinction that Hawthorne, Teitel, and others have aimed at is available, and that an adequate definition of determinism that captures something like the intuition behind Full Determinism, properly understood, is available. To see it, though, requires us to shift back to the other post-Lewisian thread, the one that was prematurely cut off by Belot (1995a) and Belot (1995b). We intend to take this line of development back up, and to argue that it is a genuine problem-solver for questions about determinism. In other words, insofar as Lewis was continuing the Carnapian program of explication, then he was on the right track.

This line of development consists of three positive papers, and two negative papers. On the positive side, Butterfield (1987; 1988; 1989) points out that there is an imprecision in Lewis’ talk of “diverging worlds”. Since, for Lewis, no individual can exist in two worlds, distinct worlds cannot overlap. Butterfield then proposes that Lewis needs a notion of a *duplication map* $g : U \rightarrow U'$, where U is an initial segment of W , and U' is an initial segment of W' . With this notion in hand, Butterfield explains that there are two precisifications of Lewis’ notion of determinism, a stronger one “DM2”, and a weaker one “DM1”. Roughly speaking, DM2 says that if there is a duplication $g : U \rightarrow U'$ of initial segments of worlds W, W' , then there is a duplication $f : W \rightarrow W'$ of these worlds.¹⁰ Then Butterfield goes on to apply this analysis to the hole argument, and shows that GR satisfies DM2.

This is an amazing outcome. Butterfield seems to have blocked the hole argument, vindicating substantivalism and Lewisian counterpart theory, not to speak of formal definitions of determinism. But then along came Gordon Belot. Belot (1995b) shows that even Butterfield’s stronger condition DM2 is too liberal by giving an example of a clearly indeterministic process that satisfies it. He then goes on to give two refined and strengthened versions of DM2, but immediately provides counterexamples to them. The upshot seems clear: do not try to turn Lewis’ metaphysical definition of determinism into a Carnapian explication, because formal definitions will never capture the

¹⁰This discussion occurred in the context of the hole argument, and so the definitions were originally stated in terms of manifolds, tensors, and smooth mappings. But the structure of Butterfield’s definitions are independent of these details. For further development in those terms, specifically for GR, see Manchak, Barrett, et al., 2025.

full sense of determinism. At least that seems to have been the lesson that many philosophers – Belot included – took away from his arguments.

We have already argued against this general posture. But we also think Belot’s arguments fail on their own terms. That is, we think that Belot himself gave a promising formal definition of determinism. We will now argue that his “counterexample” is nothing of the sort. It does *not* show the inadequacy of his precisification of Lewis’ diverging-worlds definition of determinism.

Belot’s (1995b) first definition of determinism is essentially a direct transcription of Butterfield’s DM2:

D1: A world W is deterministic if, whenever W' is physically possible with respect to W and t, t' , and $f : W_t \rightarrow W_{t'}$ are such that f is a duplication, there is some duplication $g : W \rightarrow W'$.

Belot argues that D1 does not capture determinism in its fullest sense, since there are indeterministic processes that are D1-deterministic. We agree. Consider, again, the Hawthorne’s identical ships. This example satisfies D1, but it still seems to be indeterministic in some sense.

After dismissing D1, Belot considers the following strengthened version:

D2: W is deterministic if, whenever W' is physically possible with respect to W , and t, t' , and $f : W_t \rightarrow W'_{t'}$ are such that f is a duplication, there is some duplication $g : W \rightarrow W'$ whose restriction to W_t is f .

The key difference between D1 and D2 is that the latter requires a relationship between the duplication $g : W \rightarrow W'$ and the duplication $f : W_t \rightarrow W'_{t'}$, viz. f is the restriction of g to W_t . The fact that D2 is genuinely stronger than D1 depends on the assumption that “agreement” can be witnessed by various functions. Indeed, if agreement were a binary relation on worlds (or world segments), then D1 would imply D2. This might explain why modal metaphysicians have overlooked D2 (or the even stronger version D3, that we will soon consider). Modal metaphysicians have tended to think in terms of binary relations on worlds, whereas D2 asks us to keep track of different ways that worlds can be matched with each other.

Belot then provides a counterexample to D2, i.e. an example that satisfies D2, but that seems obviously to be indeterministic. Once again, we agree

that D2 does not capture the full sense of determinism. Belot then considers one final proposal.¹¹

D3: A world W is deterministic if, whenever W' is physically possible with respect to W , and t, t', W' and $f : W_t \rightarrow W'_t$ are such that f is duplication, then there is exactly one duplication $g : W \rightarrow W'$ which extends f .

Here Belot gives yet another counterexample, and declares the prospects for a formal definition of determinism to be grim.

Some of the details of D3 are inessential, and it can easily be made into a schematic that applies to just about any scientific theory. For example, while D3 is formulated in terms of possible worlds, we will sometimes talk instead about models (of a theory). Similarly, D3 takes the determiner to be a time-slice W_t , but we can take it to be other parts of a world or a model, e.g. an initial segment of a possible world (see Lewis, 1983; Butterfield, 1989), or an initial data surface embedded in a four-dimensional Lorentzian manifold (see Landsman, 2023). The details may differ, but all of these cases conform to the following schematic:

$$\begin{array}{ccc} M & \overset{g}{\dashrightarrow} & M' \\ \uparrow i & & \uparrow i' \\ U & \xrightarrow{f} & U' \end{array}$$

Here $i : U \rightarrow M$ and $i' : U' \rightarrow M'$ are the embeddings of initial segments into the entire history, and $f : U \rightarrow U'$ is an isomorphism of initial segments. D3 then says: determinism holds just in case any isomorphism of initial segments extends uniquely to an isomorphism of worlds.

In what follows, we will argue that D3 is the best definition of determinism on the market. For now, we will simply point out that D3 holds for General Relativity.¹²

¹¹Belot's D3 seems to have been independently rediscovered by Landsman (2023) and Cudek (2024).

¹²To be more precise, D3 holds for the category of globally hyperbolic relativistic spacetimes, where morphisms are assumed to preserve the metric. (This result follows directly from the Choquet-Bruhat-Geroch theorem; see the account in (Landsman, 2023).) A more careful treatment of determinism in GR is given in Manchak, Barrett, et al., 2025. Our attitude here should be clarified by what we say in section 3 about electromagnetism. The

6 Belot against D3

We think that D3 is an excellent definition of when a theory is deterministic. But that is not the conclusion that its architect, Belot, drew. In fact, Belot's harsh criticism of D3 made it all but invisible to philosophers for thirty years, until it reappeared in work by Halvorson and Manchak (2022), Landsman (2023), and Cudek (2024). In this section, we consider Belot's purported counterexample to D3, and we argue that it is nothing of the sort.

Belot argues that D3 misclassifies the following example as deterministic.

In this example, W is a world with spacetime points and Newtonian spacetime structure. It initially contains a single α particle. The laws of nature decree that one year later, at $t = 1$, the α particle decays into continuum many β particles; arranged so that at time t , the β particles form a spherical shell of radius t ; with each β particle moving away from the center of the sphere along its radius.¹³ (Belot, 1995b, p 193)

This example certainly does seem indeterministic: given one way in which the cloud of β particles is expanding, there is another way that it could have been. But there is an ambiguity in the formulation of the underlying theory: does it permit rotations of β particles that do not also rotate spatial points? An affirmative answer leads to an indeterministic theory, and a negative answer leads to a deterministic theory. Far from undermining D3, Belot's example shows that D3 is sensitive to subtle differences of theory formulation.

Belot's argument that the example is D3-deterministic relies on the following claims:

1. As for spacetime points, duplications must preserve metric relations, hence they are symmetries of Newtonian spacetime (shifts, spatial rotations).
2. As for a material particle stage a , a duplication must preserve the relation $\ell(a) = p$ of a being located at spacetime point p .

difference between two versions of EM manifests itself in different choices of morphisms between models; and one of these two choices leads to a better theory (in our opinion, and in the opinion of most physicists).

¹³We changed the time scale for ease of exposition.

3. The only Newtonian symmetry that preserves the worldline of the α particle is a rotation around the timelike line that extends that particle's trajectory.

The upshot is clear: a symmetry of the α particle's trajectory can only be extended in one way to a symmetry of everything in the model (both spacetime points and material particles). Criterion D3 is satisfied.

Belot then argues that the example should be conceived of as indeterministic. His argument relies on the following claim:

There is a legitimate counterpart relation g' (not a duplication) that moves material particles but not spacetime points. Thus, g' breaks the relation $\ell(a) = p$ of a material particle a being located at spacetime point p .

If we entertain g' as a legitimate counterpart relation, then that is tantamount to adding additional symmetries. Now a rotation of $S_{t \leq 0}$ has more than one extension to a symmetry of the joint system of spatial points and material particles: it can either be extended to a rotation of both space and particles, or it can be extended to a rotation of just space but not particles. Therefore, the extension of symmetries is not unique, and condition D3 is violated.

For some metaphysicians, the issue at stake here might be the distinction between a "duplication" $g : W \rightarrow W'$ and a more general counterpart function $g' : W \rightarrow W'$. And perhaps Belot's point was not to argue against D3 as an analysis of determinism, but to argue against an implementation of D3 where the morphisms have to be Lewisian duplications. We do not know, and we are not convinced that we will understand scientific theories better by going deeper into this issue. What Belot's example reveals, we think, is that whether or not a theory is deterministic depends crucially on the precise formulation of the theory, where the precise formulation includes a notion of isomorphism between models. Different notions of isomorphism correspond to different theories; and one of these theories might be deterministic while the other is indeterministic.

6.1 Belot's example, two ways

In what follows, we will precisify and simplify Belot's example, so as to spare the reader from irrelevant technicalities. But what we show for these simplified theories generalizes straightforwardly to theories based on Newtonian spacetime, and with uncountably many material particles.

Newtonian spacetime has the feature that spatial points maintain their identity over time — and hence it makes sense to talk about whether an object is changing its position over time. (This in contrast to Galilean spacetime.) Newtonian spacetime also has a rather small group of symmetries: uniform shifts and rotations around vertical (timelike) lines. These features of Newtonian spacetime play an important role in the setup of Belot’s third example.

Suppose now that we want to construct an analogue of Newtonian spacetime, but with finitely many spatial locations and finitely many times. In fact, for our purposes it will suffice to have three locations and two times — an initial time $t = 0$ and a final time $t = 1$. We do not need shifts (which play no role in Belot’s example), but we will allow for rotations around the center location.

There are two ways to set up such a framework. These two ways are formally equivalent in the case of Newtonian spacetime.

1. Represent spacetime by a family of types S_t , with t a time parameter, and postulate a “persistence” relation (the analogue of an affine connection) between the types.¹⁴ The persistence relation can be represented by isomorphisms with compatibility relations. To represent Newtonian spacetime, we assume a unique isomorphism $\delta_{t,t'} : S_t \rightarrow S_{t'}$.
2. Represent space by a type S and time by another type R , so that spacetime is represented by the product type $R \times S$.

The advantage of the first, more complicated, setup is that it generalizes more easily, e.g. to Galilean spacetime. The advantage of the second setup is that we do not have to keep track of sorts. Since Belot’s example assumes Newtonian spacetime, we will start with the second approach.

Let Σ be a signature with two sort symbols S for spatial points. We could then add the axiom that there are three things of sort S , corresponding to the three particle positions. But it is simpler just to ignore that α particle, which effectively defines a constant symbol (a name for its location). Thus we take as our first axiom:

There are two things of sort S .

¹⁴Here we use “type” and “sort” synonymously, both in the sense of many-sorted logic (see Halvorson, 2019). We are using this framework for its flexibility, and without any commitment to type theory as a foundation of mathematics.

β_1		β_2
	α	

Figure 1: Space with three places and two times, with two β particles at the later time. Since symmetries are assumed to fix α , the center blocks can be omitted from the model without changing the conclusions we draw.

We now have a non-trivial decision to make about how to represent the β particles. It would seem natural to assume that β particles belong to the ontology of the theory; and hence that we should introduce a new sort symbol B for them. Thus we add a second axiom:

There are two things of sort B .

and a function symbol ℓ to indicate where each β particle is located at time $t = 1$. Let's call this theory T_q .

The theory T_q is actually quite trivial: it's bi-interpretable with the theory (in empty signature) that says there are exactly two things. A model M of T_q consists of two sets S^M and B^M , each with two elements, and an isomorphism $\ell^M : B \rightarrow S$. For any two models M, M' of T_q , there are two isomorphisms $h, k : M \rightarrow M'$. (As T_q has two sort symbols S and B , a homomorphism $h : M \rightarrow M'$ consists of two maps $h_S : S^M \rightarrow S^{M'}$ and $h_B : B^M \rightarrow B^{M'}$, such that $\ell^{M'} \circ h_B = h_S \circ \ell^M$.) When $M' = M$, the cross-model identity relation picks out one of the two isomorphisms $h, k : M \rightarrow M'$ as the identity 1_M , while the other is the flip π .

Note that T_q does not assign any distinguishing properties to β particles. While we might be tempted to ask, "couldn't that β particle be located at a different place than it is?" the language of T_q does not have any expression "that β particle". We might also be tempted to count possibilities: given one pairing of β particles and spatial locations, surely there is a second, distinct pairing? But the language of T_q does not recognize "pairings" as countable objects. In fact, since T_q says that there is an isomorphism between B and

S , it could be interpreted “relationally” as saying that the identity of spatial locations is uniquely determined by the β particles that occupy them — and so it makes no sense to consider moving a β particle to a different spatial location.

Belot argues that D3 wrongly judges this example to be deterministic. However, when the example is precisified in this way, i.e. as theory T_q , we contend that it is deterministic. The initial condition is: space is empty. (The α particle is irrelevant, as it simply served to restrict symmetries of space to rotations.) The final condition is: there are two β particles, one in each location. But here ordinary language (supplemented by possible-worlds talk) misleads us into thinking that there is *one* initial condition that could lead to either one of *two* final conditions. In fact, there is no sense, according to theory T_q , in which there are more final conditions than initial conditions.

To be precise, consider two embeddings of initial conditions $i : S \rightarrow M$ and $i' : S' \rightarrow M'$. Here S and S' are sets with two elements, and M (respectively M') adds a second set B (respectively B') and an isomorphism $\ell : B \rightarrow S$ (respectively $\ell' : B' \rightarrow S'$). In this precisification, the β particles do not have any “whichness” that the spatial points lack. Hence, any isomorphism $f : S \rightarrow S'$ of space is compatible with precisely one isomorphism $\ell' \circ f \circ \ell$ of β particles:

$$\begin{array}{ccc} B & \dashrightarrow & B' \\ \ell \uparrow & & \uparrow \ell' \\ S & \xrightarrow{f} & S' \end{array}$$

Therefore, T_q is D3-deterministic.

We are not saying that T_q is the one and only way of understanding Belot’s example. To the contrary, we are saying that T_q is one way of understanding Belot’s example to be a deterministic theory. If one has the intuition — as Belot did — that there is another sense in which the example is indeterministic, then, we claim, that is because one has a different theory in mind. In particular, if one considers that “being occupied by a particular β particle” is a property of spatial points, then Belot’s example is D3-indeterministic.

To be precise, let T_h be a theory with a single sort symbol S , and with a name b for one of the two β particles. We consider $\lceil x = b \rceil$ to represent the property that the spatial point x is occupied by b at $t = 1$. Thus, a model M of T_h consists of a set S with two elements and a distinguished element

$b^M \in S$, and an isomorphism $g : M \rightarrow M'$ consists of a function from S to S' such that $g(b^M) = b^{M'}$. Thus, there is a unique isomorphism between any two models of T_h .

For the theory T_h , an initial condition is a set S with two elements, and a final condition is the same set S plus the choice of one of the two elements $b^M \in S$. This choice leads to a reduction of symmetry — and hence to a breakdown of determinism. Indeed, consider the isomorphism $f = 1_S : S \rightarrow S$ of initial conditions. Let $b^M \in S$, and let $b^{M'}$ be the other element of S . Then there is no isomorphism $g : M \rightarrow M'$ that completes the following diagram:

$$\begin{array}{ccc} M & \overset{g}{\dashrightarrow} & M' \\ \uparrow i & & \uparrow i' \\ S & \xrightarrow{f} & S \end{array}$$

Therefore, T_h is D3-indeterministic.

Let's take stock. The theories T_q and T_h seem to describe the same worlds: at the initial time there are two places and no material particles; and at a subsequent time, each place is occupied by a β particle. However, there is a subtle difference between the two descriptions: in T_h , the β particles have names. In contrast, T_q does not name the β particles, and so any symmetry of spacetime points can be lifted to a corresponding symmetry of β particles.

What we just said about names (in the “formal mode”) can be rephrased as a statement about properties (in the “material mode”). The theory T_q sees no properties that distinguish spacetime points from each other, and no properties that distinguish β particles from each other. Thus, for T_q , the domain of particles and the domain of spacetime points have the same amount of structure. In contrast, T_h has a property $\ulcorner x = \beta \urcorner$ that breaks the symmetry between the two spatial points. Hence, a space that was originally rotationally symmetric enters into a state that is not rotationally symmetric. This is why T_h is indeterministic: symmetry is broken.

Informally described examples can be helpful intuition pumps, but it is all too easy to introduce ambiguities. In the original example, Belot did not provide a precise theory of material particles and their relations to spacetime points. According to one precisification, T_q , material particles are indistinguishable, and symmetries of spacetime extend uniquely to symmetries of material particles. In this case, Belot's example is unambiguously determin-

istic. According to another precisification, T_h , material particles do have distinguishing properties. In this case, a symmetry of empty space at one time does not necessarily extend to a symmetry of space plus material contents at a later time. In this case, Belot’s example is unambiguously indeterministic.

We conclude this section on an ironical note: Belot claims that determinism is not a “formal property of uninterpreted theories.” But his informal example is neither deterministic nor indeterministic until it is formulated in a precise manner, i.e. by transforming it into a formal theory.

6.2 Failure of uniqueness

Belot’s condition D3 requires a unique extension for each isomorphism of initial conditions. Consequently, this condition can fail in one of two ways: (i) an isomorphism of initial conditions does not extend to an isomorphism of models, and (ii) an isomorphism of initial conditions extends in multiple ways to an isomorphism of models. We have seen how the first condition can fail, and now we will see how the second condition can fail.

Suppose that at the initial time there is an α particle, and it then spawns twin β particles. (Here the assumption that the β particles are twins is required to draw the conclusion that the theory is indeterministic.) Crucially, we do *not* assume that there are spatial points S in addition to the α and β particles. In Belot’s example, while the α particle only exists at the initial time, and the β particles only exist at the final time, the background space S exists at all times, and it prevents us from mixing and matching symmetries of α and β particles.

In our current example, there are no atemporal spatial points S , and so no compatibility condition between symmetries of α particles and symmetries of β particles. Hence, D3-determinism fails: the unique (identity) symmetry of the α particle can be extended to two distinct symmetries of β particles, either the identity or the flip. (If your intuition is that this example is deterministic, then you might be imagining that all these particles are located in space, so that the initial state does have non-trivial symmetries.)

This example shows that D3 is sensitive to what metaphysicians have suggested are “merely haecceitistic differences”, but without mentioning haecceities, and without having names for objects. The theory we have described (call it \hat{T}_q) permits only one possible world, viz. a world in which there is one α particle at $t = 0$, and two β particles at $t = 1$. It may be tempting to say that a haecceitistic interpretation of \hat{T}_q would consider there to be

two worlds: an original world W , and a world W' in which the two β particles have been flipped. But that is not a legitimate interpretation of \hat{T}_q , since “flipping β particles” does not witness the existence of non-isomorphic models.

There is a nearby theory that could be said to be haecceitistic. Let b be the name for one of the two β particles, and suppose that there are two properties, say p and $\neg p$. Then this theory \hat{T}_h has two non-isomorphic models: one in which $p(b)$ and one in which $\neg p(b)$. This theory is also D3-indeterministic.

6.3 The difference a spacetime can make

In Belot’s example, it seems clear that things could have turned out differently than they did, i.e. the roles of β_1 and β_2 could have been reversed. But when we asked “what does ‘differently’ mean here?”, we had to mention absolute space S , and various ways that a name could be attached to points of this absolute space. In other words, the argument that T_h is indeterministic depends crucially on the fact that the points of space already exist prior to the emergence of the β particles, and so they provide a standard by which we can count the number of possibilities.

These considerations suggest that the intuition that T_h is indeterministic might depend on the assumption of absolute space. Here we show that this is correct: in the absence of the assumption of absolute space, the theory with names for β particles is deterministic, while the theory without names for β particles is indeterministic.

Newtonian spacetime has a lot of structure, and relatively few symmetries. This lack of symmetries plays a crucial role in Belot’s argument: the only Newtonian symmetries that preserve that α particle trajectory are rotations. That constraint would not hold in a spacetime that lacks inertial structure, such as Leibnizian spacetime. The key structural difference between Newtonian and Leibnizian spacetimes is the following (see Earman, 1989; Weatherall, 2021): in Newtonian spacetime N , a symmetry’s action on a past segment $\{N : t \leq 0\}$ determines that symmetry’s action on all of N . In Leibnizian spacetime $M \cong \mathbb{R} \times \mathbb{R}^3$, a symmetry’s action on a past segment does *not* determine its action on all of M . For example, for $t \leq 0$, let σ_t be the identity map on \mathbb{R}^3 ; and for $t > 0$, let $\sigma_t(x) = x + tv$, where v is a fixed non-zero vector in \mathbb{R}^3 . Then the family $\{\sigma_t : t \in \mathbb{R}\}$ defines a symmetry of Leibnizian spacetime.

To mimic the flexibility of Leibnizian spacetime, we can use two sorts S_0 and S_1 to represent space at the different times. We can again represent the α particle as a constant symbol of sort S_0 ; but this particle's existence makes no difference to the example, since a constraint on the symmetries of S_0 imposes no constraint on the symmetries of S_1 . We also must choose again between representing β particles by a domain of indistinguishable entities, or by a domain of distinguishable entities. Let \check{T}_q be the theory with an additional sort symbol B , and a function symbol $\ell : B \rightarrow S_1$ which is assumed to be an isomorphism. The astute reader will see that adding B and an isomorphism is structurally equivalent to doing nothing at all, so we might as well simply omit B . Let \check{T}_h be the theory with an additional constant symbol b of sort S_1 .

It is easy to see that \check{T}_q is D3-indeterministic. A model M of \check{T}_q consists of two sets S_0, S_1 , each of which has two elements. Thus, any pair of bijections $f_0 : S_0 \rightarrow S'_0$ and $f_1 : S_1 \rightarrow S'_1$ determines an isomorphism $\langle f_1, f_2 \rangle : M \rightarrow M'$. Since there are two distinct automorphisms of S_1 , there are, for any model M of \check{T}_q , two distinct automorphisms of M that restrict to the identity automorphism on initial conditions. Therefore, \check{T}_q is indeterministic, in contrast to the version of this theory that is set in Newtonian spacetime. Intuitively the reason that \check{T}_q is indeterministic is because there simply is no connection between the state at $t = 0$ and the state at $t = 1$. Notice that it is completely irrelevant whether or not we are haecceitists, or how we count the number of initial and final states. The only relevant point is how symmetries of initial states are connected (or not) to symmetries of final states.

It is also easy to see that \check{T}_h is D3-deterministic. Since there is simply no connection between initial and final conditions, the extendability condition is fulfilled. Furthermore, for any two final conditions S_1 and S'_1 , there is a unique isomorphism $f_1 : S_1 \rightarrow S'_1$, since f_1 must map b^M to $b^{M'}$. Thus, each isomorphism between initial conditions extends uniquely to an isomorphism of final conditions, and \check{T}_h is deterministic.

The verdict about which theory is deterministic (\check{T}_q or \check{T}_h) is flipped in the passage from Newtonian to Leibnizian spacetime. This result shows that it would be a mistake to think that the reason that the Newtonian theory, T_h , is indeterministic is because it is a "haecceitistic" theory, with names for individuals. In terms of β particles, the Leibnizian theory, \check{T}_h , is just as "haecceitistic" as T_h . And in terms of spacetime points, neither \check{T}_h nor T_h has names for them, while both can indirectly name them by saying whether they are occupied by particular β particles. On reflection, this is how it should

be. The judgment of determinism has nothing to do with the atemporal thesis of haecceitism, and everything to do with the relations that the theory postulates between objects at different times.

7 Bridging a non-existing gap

Teitel (2019) argues that metaphysicians have an important job in uncovering what modal-metaphysical commitments might be required to maintain the consistency of spacetime substantivalism with full determinism. He poses the challenge as “bridging the gap” between qualitative and full determinism.

We need a doctrine that . . . bridges the crucial gap between GR’s qualitative determinism and its full determinism (thereby resolving both the original hole argument and my revised hole argument). (Teitel, 2019, p 379)

Any of those three anti-haecceitistic doctrines suffices to bridge the gap between GR’s qualitative and full determinism. (Teitel, 2019, pp 359-360)

I deliberately set up the issues surrounding the hole argument by directly discussing modality and which doctrines imply the right modal correlations to bridge GR’s qualitative and full determinism, rather than following the standard practice in the literature of theorizing primarily in terms of mathematical solution spaces and what we use them to represent. (Teitel, 2019, p 388)

We agree that if there were a gap between the sense in which GR is deterministic and some more metaphysically significant kind of determinism, then it would be worth inquiring into what metaphysical commitments are needed to bridge this gap. We claim, however, that there is no such gap.¹⁵

The hole argument raises many technical issues that are beyond the scope of this paper. Fortunately, the literature of the past thirty years offers numerous toy examples that are supposed to be analogous to GR in being qualitatively, but not fully, deterministic. (See Figure 7). We have encountered two

¹⁵Of course, this does not mean we do not recognize different senses, or “strengths”, of determinism. For instance, Belot’s D1 and D2 are weaker than our preferred D3, and may well be viewed as capturing senses of “qualitative determinism”. The crucial point is that GR is deterministic in a stronger sense than either of these, and so there is no gap for determinism in GR.

already: Belot's β particles and Hawthorne's ships. The doubly-symmetric world described by Melia (1999) provides yet another. We will now show how Hawthorne and Melia's examples, like Belot's, illustrate the distinction between D1 and D3.

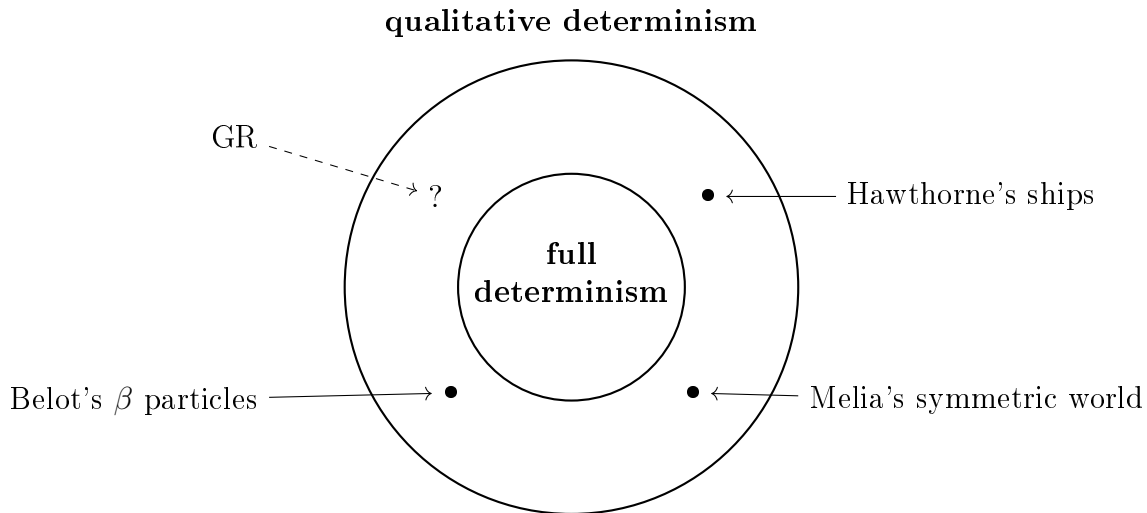


Figure 2: Theories that are supposed to be qualitatively, but not fully, deterministic.

As a warmup, consider one of Melia's simpler (but more entertaining) examples:

We could imagine a collection of bald philosophers, sitting in a circle. It is a law that one of them will grow a single hair. But, by the symmetry of the situation, *any* of the philosophers could be the lucky one. Again, our intuition is that there are many qualitatively isomorphic but distinct possibilities, each representing a different way in which the situation could evolve. (Melia, 1999, p 650)

Intuitively this example is qualitatively deterministic, since any two possible final conditions are qualitatively identical. And yet, this example clearly isn't fully deterministic, since the law does not stipulate *which* philosopher will grow a hair. We agree that this example is deterministic in one sense, but

not another. Only we think that the correct analysis is that this example is D1 but not-D3.

To see this, suppose that there are initially $n > 1$ philosophers, and that p_t represents the property of being bald at time t . Let T be the theory with axioms $\forall x p_0(x)$ and $\exists! x \neg p_1(x)$. Since T entails that p_0 holds of all things, the predicate symbol p_0 plays no role in the analysis, and we may drop it. If we set $p = p_1$ for notational simplicity, then a model M of T is determined by a set S and a singleton subset $p^M \subseteq S$. If M, M' are models of T with the same initial conditions (i.e. the same domain S), then there is at least one bijection $g : S \rightarrow S'$ such that $g(p^M) = p^{M'}$. Thus, $g : M \rightarrow M'$ is an isomorphism, and D1 is automatically satisfied.

We now show that D3 fails. Let M be a model of T whose domain S has two elements; and let M' be the model that has the same domain as M , but where the extension p^M has been switched to the other element of the domain, i.e. $p^M \neq p^{M'}$. (Note that M and M' are isomorphic models.) Then the identity 1_S is an isomorphism between the initial conditions of M and M' . However, if 1_S were an isomorphism of M to M' , then it would follow that $p^M = p^{M'}$, contradicting the definition of M' . Therefore, T is D3-indeterministic.

Melia's bald philosophers example is supposed to be a paradigm case where qualitative, but not full, determinism holds. But our split intuitions about this example can be explained by a more clear distinction, viz. that between D1 and D3. The bald philosophers example does not provide any support for the legend that there is a deeper, metaphysical sense of determinism that cannot be captured by a formal definition.

But Melia has another trick up his sleeve: an example so clever in conception that one feels sure that the quest for a formal definition of determinism will have to be abandoned.

Consider a world whose initial conditions consist of the following situation (see Figure 3). The two white particles are duplicates of each other and the two black particles are duplicates of each other. The laws in this world dictate that, after a certain fixed period of time, each black particle will start moving at a fixed velocity in a straight line towards a white particle, and that the two black particles will move towards *different* white particles. Using names for the objects found in the situation above, after a fixed amount of time either **c** will head towards **b** and **d** will

head towards **a**, or **c** will head towards **a** and **d** will head towards **b**. (Melia, 1999, p 661)

Once again, it is clear that the two possible final conditions are qualitatively identical, and hence that this example is qualitatively deterministic. But surely, one thinks, there is a *haecceitistic* difference between the two final conditions, and so the example is not fully deterministic.

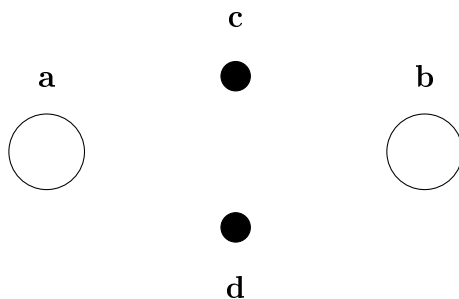


Figure 3: Melia's symmetric world

We will show, once again, that what this example illustrates is the distinction between the definitions D1 and D3. Here we just have to be a bit more careful with how we explicate the details of the example.

The most straightforward regimentation of Melia's example has two domains W and B for the particles, and for each $t = 0, 1$, a relation $\alpha_t(x, y)$ to indicate that at time t , x is adjacent to y . We add the axioms that at $t = 0$, no two things are adjacent, and that at $t = 1$, the adjacency relation induces a bijection between W and B .¹⁶ Since T dictates that α_0 is empty, its presence is structurally irrelevant. Thus we drop α_0 and set $\alpha = \alpha_1$. Let's call the resulting theory T .

A model M of T consists of two sets, each with two elements, and a bijection α^M between them. Here the initial conditions are just the two sets, whereas the final conditions include the bijection α^M . It is clear that any two models M, M' of T are isomorphic, and so T automatically satisfies D1.

It will now be easy to see that T does *not* satisfy D3, and the reason is that the final conditions have less symmetry than the initial conditions.

¹⁶We can imagine the white and black particles as sitting on the vertices of a kite, and we can think of α as a metric, where at $t = 0$, each white particle is distance $\sqrt{5}$ from each black particles; and at $t = 1$, each white particle is distance 0 from one black particle, and distance 4 from the other.

More rigorously, in a model M of T , the initial conditions (i.e. the two sets W and B) are invariant under any symmetry of the form $\langle f_0, f_1 \rangle$, where $f_0 : W \rightarrow W$ and $f_1 : B \rightarrow B$ are bijections. In contrast, the final conditions include a bijection $\alpha^M : W \rightarrow B$, and this bijection is not invariant under symmetries of the form $\langle f_0, f_1 \rangle$ where f_0 and f_1 do not have the same polarity (i.e. where f_0 is the identity and f_1 flips elements or vice versa). Therefore, there is a symmetry of initial conditions that does not extend to a symmetry of models, and T is D3-indeterministic.

The upshot: to understand the sense in which Melia’s doubly-symmetric world is indeterministic, we do not need to know anything about haecceitistic differences. It is enough to see that there is a duplication of initial conditions that does not extend to a duplication of worlds. So this example, and others like it, only emphasizes the virtues of “purely formal” definitions of determinism.

Similar remarks can be made about Hawthorne’s ships. As with Belot’s and Melia’s examples, his example is D1 but not D3 deterministic — reinforcing our claim that this distinction suffices to capture our intuitions.

To be precise, let T be the theory with a single unary predicate p that says: there are exactly two things, and one of them is p . Here we take $p(x)$ to mean that x sinks at $t = 1$. This theory is as simple as can be imagined. A model M of T consists of a set with two elements and a singleton extension for p . For any two models M and M' of T , there is a unique isomorphism $f : M \rightarrow M'$. This shows that T is D1-deterministic.

But T is not D3 deterministic. To see this, let $S = S' = \{a, b\}$, let $i : S \rightarrow M$ be the embedding into a model M such that $M \models p(a)$ [i.e., a sinks], and let $i' : S' \rightarrow M'$ be the embedding into a model M' such that $M' \models p(b)$ [i.e., b sinks]. Then $1_S : S \rightarrow S'$ is an isomorphism of initial conditions, but there is no $g : M \rightarrow M'$ that completes the following diagram:

$$\begin{array}{ccc} M & \overset{g}{\dashrightarrow} & M' \\ i \uparrow & & \uparrow i' \\ S & \xrightarrow{1_S} & S' \end{array}$$

Therefore, T is D3-indeterministic.

In summary, the examples by Belot, Melia, and Hawthorne have been thought to provide evidence for the existence of a gap between qualitative and full determinism. However, these examples sooner illustrate the distinction between D1 and D3.

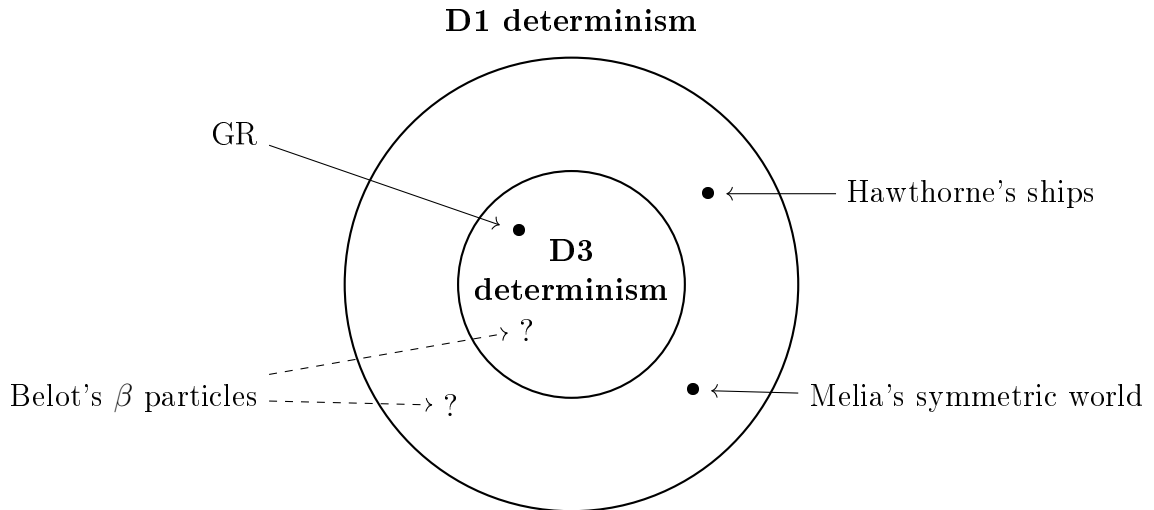


Figure 4: The toy examples in the literature are D3-indeterministic, while GR is D3-deterministic. Belot's example can be interpreted in two ways, one deterministic and one indeterministic.

8 D3 versus Full Determinism

We have suggested that D1 captures the concept of qualitative determinism, and the literature agrees with us on this. Where intuitions might still clash is whether D3 captures the strongest sense of determinism that it makes sense to ask about. We claim that it does. We will support this claim by comparing D3 with FD.

Doing so is tricky, because D3 is applicable to any theory that has a distinction between initial segments and full models. In contrast, FD assumes an additional distinction between a de re language Σ_{dr} and a qualitative sub-language Σ_q . (Actually, the definition of FD does not assume such a distinction; but as we argued above, without such a distinction, there will be no contrasting notion of qualitative determinism.) We will begin by considering the special case of theories that already have sufficiently many names. In that case, we have the nice result that FD, D1, D2, and D3 are all equivalent. But this result also shows why it is important to look at theories *without* names, where FD does not apply, because only those theories illustrate the distinctions between D1, D2, and D3.

To make things more clear, we remind the reader of a simple point: a

Σ_{dr} -theory T is fully deterministic iff T satisfies D1.

Proposition 2. *For theories with sufficiently many names, FD, D1, D2, and D3 are equivalent.*

Proof. As noted above, the definition of FD is formulated against a background assumption that there are two signatures $\Sigma_{\text{q}} \subset \Sigma_{\text{dr}}$. However, FD is just D1 for a Σ_{dr} -theory, and we are assuming that T is such a theory. It will suffice then to show that D1 implies D3. Suppose then that T satisfies D1, and let $f : U \rightarrow U'$ be an isomorphism. By D1, there is an isomorphism $g : M \rightarrow M'$. Since all elements of U and U' are named, the isomorphism $f : U \rightarrow U'$ is unique, hence $g|_U = f$. That is, g extends f . Similarly, since all elements of M and M' are named, the isomorphism $g : M \rightarrow M'$ is unique. Therefore g is the unique extension of f , and T satisfies D3. \square

But what about the case where we do not have names? One natural suggestion for extending FD to such theories is simply to add names when necessary. Given a Σ -theory T , let Σ^+ be the expansion of Σ to include “sufficiently many” names, and let T^+ be the extension of T to Σ^+ . We can then ask about the relation between T having property D2 or D3 and T^+ having property FD (i.e., D1). In some cases, D2 (and D3) for T do imply FD for the enriched theory. In particular, this happens for theories of enduring objects – that is, theories whose models’ domains coincide with the domains of their initial segments. D2 for a theory T implies that the name-enriched theory T^+ satisfies FD.

Proposition 3. *Suppose that T is a theory of enduring objects. If T satisfies D2, then T^+ satisfies FD.*

Proof. Suppose that T satisfies D2. Let M, M' be models of T^+ with initial segments U, U' such that $U \sim_{\text{dr}} U'$. That is, there is a Σ^+ -isomorphism $f : U \rightarrow U'$. Since T and T^+ are theories of enduring objects, $|M| = |U|$ and $|M'| = |U'|$. It follows that $f(c^M) = c^{M'}$, for all names c . Since f is a Σ -isomorphism, D2 entails that there is a Σ -isomorphism $g : M|_{\Sigma} \rightarrow M'|_{\Sigma}$ that extends f . Since g extends f , $g(c^M) = f(c^M) = c^{M'}$, for all names c . Therefore, g is a Σ^+ -isomorphism, and $M \sim_{\text{dr}} M'$. Therefore, T^+ satisfies FD. \square

What drives this result is that if objects persists over time, then fixing the referents of constant symbols in initial segments fixes those referents in

the entire model. But as we have seen, not all theories have enduring objects. In general, we find that even D3 for a theory T comes apart from FD for an enriched theory T^+ . First we now show that T^+ satisfying FD does not imply that T satisfies D3.

Example (toy Leibnizian spacetime). Let T be a theory with two sort symbols S_0, S_1 , and axioms that say that there are exactly two objects of each sort. The theory T satisfies D2, but not D3, since an automorphism of the first sort S_0 extends in more than one way to an automorphism of the second sort S_1 . The theory T^+ adds two constant symbols of each sort, say a_0, b_0 and a_1, b_1 . But then for any models M, M' of T^+ , there is an isomorphism $f : M \rightarrow M'$, and so $M \sim_{\text{dr}} M'$. Therefore, T^+ satisfies FD. \square

The converse also fails: that is, a theory T may satisfy D3, but the enriched theory T^+ with sufficiently many names might not satisfy FD.

Example (toy Newtonian spacetime). We now consider Newtonian spacetime as having a different domain S_t of spatial points for each time, and isomorphisms $\delta_{t,t'} : S_t \rightarrow S_{t'}$ that pick out the preferred frame of reference. For our purposes, it suffices to consider a simple case with two sorts S_0, S_1 , and a single function $\delta : S_0 \rightarrow S_1$. Let T be the theory that says there are exactly two elements of type S_0 , and that δ is a bijection. If M, M' are models of T , then an isomorphism $g : M \rightarrow M'$ consists of two bijections $g_0 : S_0 \rightarrow S'_0$ and $g_1 : S_1 \rightarrow S'_1$ that satisfy the compatibility condition $\delta^{M'} \circ g_0 = g_1 \circ \delta^M$. It follows that any bijection $g_0 : S_0 \rightarrow S'_0$ extends uniquely to a bijection $g : M \rightarrow M'$. Therefore, T satisfies D3.

For the name-enriched theory T^+ , suppose that a_0, b_0 are constant symbols of sort S_0 , and that a_1, b_1 are constant symbols of sort S_1 . Then there is one model M of T^+ such that $M \models \delta(a_0) = a_1$, and a non-isomorphic model M' of T^+ such that $M' \models \delta(a_0) = b_1$. But the initial segments of M and M' are isomorphic. Therefore, T^+ does not satisfy FD. \square

Note that in the toy Newtonian example, T^+ is actually D3-indeterministic (in addition to not being fully deterministic). Therefore, adding names to a D3 deterministic theory can result in a D3 indeterministic theory. It might seem like a strike against D3 that it is not stable under the addition of names to a theory. But that intuition is based on a false assumption that the role of names in formal theories is the same as the role of names in ordinary language. In a formal theory, introducing a new name is tantamount to introducing a new property $\varphi(x) \equiv (x = c)$. But we should expect that adding

new properties, without adding dynamical laws that govern the behavior of those properties, could transform a deterministic theory into an indeterministic theory.

Something similar happens with GR. As usually formulated, GR satisfies D3 (Halvorson and Manchak, 2022). But adding names to GR results in a theory that does not satisfy D3, and thus is not fully deterministic – as shown by the hole argument. Indeed, Weatherall (2018) suggests that one way of understanding manifold substantivalism, as described by Earman and Norton (1987), is as a view on which there are additional singular, haecceitistic facts about spacetime points that can only be described by something like enriched – or, in Pooley’s terms, substantivalist – GR. The hole argument then shows this theory is indeterministic, even on D1. Of course, the important point is that this enriched theory is not GR, the theory that we have good reasons to believe at least approximately describes the structure of space and time in our universe, but rather GR plus a great deal more structure. Without that structure, GR is deterministic by D3, and FD does not even apply.

9 Interpretation revisited

Some readers may find the analysis at the end of the previous section unsatisfactory. One might argue, for instance, that we have simply misunderstood full determinism. What motivates full determinism is the idea that there are objects in the world, and a deterministic theory ought to be able to assign them properties in an unambiguous (deterministic) way. The names just give us a way of referring to those objects. When we say that FD simply does not apply to GR, that is not a problem for FD, it is a problem for GR! We need enriched GR to accurately assess whether the theory determines the properties of individuals. Without names, we are simply dwelling in the domain of the qualitative.

We think this posture is wrongheaded. But rather than argue against it directly, we want to propose a diagnosis of where it originates, drawing on the arguments from section 3. As we argued there, interpretation is itself formal. All anyone is doing when they try to interpret theories is just layering models on models. Crucially, for the present point: Tarskian semantics involves mapping theories, with or without names, into set theory, typically understood as a theory with names (or rather, as theory whose membership properties allow us to individuate sets).

We suggest that the motivation for FD arises because when you interpret a theory without names in set theory, without paying careful attention to how the semantics works, it looks as if the “real” points, the ones the theory is referring to, have names. This apparently means we can ask about what “determines” what properties those named things have. But this instinct is a mistake. It illegitimately mixes two different things: the theory we are trying to analyze, and the formal tools we use to analyze it. Determinism for theories is about whether initial segments of models determine the entire model. The “determination” of what which objects in a model carry which properties (or names) is about an interpretation map, in the Tarskian sense. In other words, failures of FD are about *us*, that is, about how we think about our formal semantics and how we define interpretation maps, not about our theories or the world.

We suggest that something similar happens in many discussions of the hole argument. Philosophers apparently mistake GR for enriched GR when thinking about substantivalism. Doing this is not only a mistake, it quickly leads to incoherence. We can see this point most starkly by considering a special sector of GR, consisting of four-dimensional, vacuum, inextendible, globally hyperbolic “Heraclitus” spacetimes (Manchak and Barrett, 2023; Manchak, Barrett, et al., 2025). Spacetimes with that property are such that every point is uniquely specified by its metrical properties (including derivatives, i.e., curvature scalars). Call the theory of Heraclitus spacetimes HGR. HGR has names, in the sense that one can uniquely refer to points. This theory satisfies FD. (Of course, it also satisfies D3.)

Now consider what happens when we interpret HGR in set theory. We assign those named points to sets, which also have names. But of course, nothing in the theory can determine which (named) set we assign to which named point. No theory can do that, because it happens at the level of choosing an interpretation map! What this means, though, is that on the set theory side we now have *too many names*. That is, if we try to doubly-interpret this theory, as we suggest the FD advocate would wish to, and run our analysis of determinism on those doubly-interpreted structures, we will find that they are not deterministic. We claim this is a completely generic situation that arises from layering interpretation on top of set theory.

And it gets worse! Suppose we somehow solved this problem in the doubly-interpreted theory, perhaps by adding laws that coordinate between the two types of names, restoring FD. What then? Now we have a new theory, with lots of redundant names, that has no expressive resources beyond our

original theory. (Perhaps the theories are even logically equivalent, depending on the details.) But then we can interpret that theory, using Tarskian semantics. The (triply) interpreted theory will now have three types of names – the two coordinated ones in our theory, plus the names of the sets on the semantic side. The problem will arise again. And so on ad infinitum. But once we see how this works, it is clearly a pseudo-problem, one arising only because of a confusion about what is part of the theory and what is not, what the theory *should* determine and what is merely structure added in interpretation.

10 Conclusion

In a spirit of methodological tolerance, we are happy to live in a world where some philosophers look more like mathematicians, and other philosophers look more like poets. But we find it odd that some prominent analytic philosophers argue against formal approaches, saying things like, “the purely formal approach is a nonstarter”, or, “determinism cannot be a formal property of theories.” The descriptive content of such claims is opaque (what is a non-formal property?), but their tacit normative content is to recommend against the very methods that distinguished analytic philosophy from the more speculative, and less science-friendly, approaches of the nineteenth century. Surely this belongs among the ironies of intellectual history.

We are also motivated by a practical concern about how to facilitate fruitful dialogue between philosophy and the natural sciences. If philosophers insist on making distinctions that cannot gain any traction in scientific practice, then they will only reinforce disciplinary boundaries that are harmful to both philosophy and the sciences.

To be clear, we are *not* arguing for a kind of science-deference that says, “if scientists don’t regularly make that distinction, then neither should philosophers”. We recognize that scientists might have practical reasons to blur over distinctions of genuine metaphysical significance. One might have thought that this is the case for qualitative and full determinism, but our investigation shows there really is no conceptual joint here. In contrast, there is a clear distinction between the three formal criteria D1, D2, and D3, and these criteria give clear answers about whether real-life scientific theories are deterministic. With this kind of division of labor, philosophy and the natural sciences can work together to figure out whether we live in a deterministic

world.

Acknowledgments

Thanks to Gordon Belot, Klaas Landsman, and Trevor Teitel for comments on a previous draft. JOW: This material is based upon work supported by the National Science Foundation under Grant No. 2419967.

References

- Belot, Gordon (1995a). “Determinism and ontology”. In: *International Studies in the Philosophy of Science* 9.1, pp. 85–101. DOI: 10.1080/02698599508573508.
- (1995b). “New work for counterpart theorists: Determinism”. In: *The British journal for the philosophy of science* 46.2, pp. 185–195. DOI: 10.1093/bjps/46.2.185.
- Bradley, C and JO Weatherall (2022). “Mathematical responses to the hole argument: Then and now”. In: *Philosophy of Science* 89.5, pp. 1223–1232. DOI: 10.1017/psa.2022.58.
- Butterfield, Jeremy (1987). “Substantivalism and determinism”. In: *International studies in the philosophy of science* 2.1, pp. 10–32. DOI: 10.1080/02698598708573300.
- (1988). “Albert Einstein meets David Lewis”. In: *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1988, pp. 65–81.
- (1989). “The hole truth”. In: *The British journal for the philosophy of science* 40.1, pp. 1–28. DOI: 10.1093/bjps/40.1.1.
- Carnap, Rudolf (1937). *Logical Syntax of Language*. Routledge.
- Coffey, Kevin (2014). “Theoretical equivalence as interpretative equivalence”. In: *The British journal for the philosophy of science* 65, pp. 821–844. DOI: 10.1093/bjps/axt034.
- Cudek, Franciszek (2024). “Counterparts, determinism, and the hole argument”. In: *The British journal for the philosophy of science*. DOI: 10.1086/729767.
- Dewar, Neil (2016). “Symmetries in physics, metaphysics, and logic”. PhD thesis. University of Oxford.

- (2023). “Interpretation and Equivalence; or, Equivalence and Interpretation”. In: *Synthese* 201, p. 119.
- (2024). “The hole argument and determinism(s)”. Unpublished manuscript. URL: <https://philsci-archive.pitt.edu/24432/>.
- Earman, John (1986). *A primer on determinism*. Springer Science & Business Media.
- (1989). *World enough and spacetime*. The MIT Press.
- Earman, John and John Norton (1987). “What Price Spacetime Substantivalism? The Hole Story”. In: *The British Journal for the Philosophy of Science* 38.4, pp. 515–525.
- Halvorson, HP (2019). *The logic in philosophy of science*. Cambridge University Press.
- Halvorson, HP and JB Manchak (2022). “Closing the hole argument”. In: *The British journal for the philosophy of science*. DOI: 10.1086/719193.
- Hawthorne, John (2006). “Determinism De Re”. In: *Metaphysical Essays*. Oxford University Press, pp. 239–243.
- Hodges, Wilfrid (1993). *Model Theory*. Cambridge University Press.
- Landsman, Klaas (2023). “Reopening the hole argument”. In: *Philosophy of Physics* 1, pp. 1–24. DOI: 10.31389/pop.12.
- Leitgeb, Hannes and André Carus (2020). “Rudolf Carnap”. In: *Stanford Encyclopedia of Philosophy*. URL: <https://plato.stanford.edu/entries/carnap>.
- Lewis, David (1983). “New work for a theory of universals”. In: *Australasian journal of philosophy* 61.4, pp. 343–377. DOI: 10.1080/00048408312341131.
- Manchak, JB and TW Barrett (2023). “A Hierarchy of Spacetime Symmetries: Holes to Heraclitus”. In: *The British Journal for the Philosophy of Science*. Forthcoming.
- Manchak, JB, TW Barrett, et al. (2025). “Determinism and Asymmetry in General Relativity”. Unpublished manuscript.
- Melia, Joseph (1999). “Holes, haecceitism and two conceptions of determinism”. In: *The British journal for the philosophy of science* 50.4, pp. 639–664. DOI: 10.1093/bjps/50.4.639.
- Montague, Richard (1962). “Deterministic theories”. In: *Decisions, Values and Groups*. Pergamon Press, pp. 325–370.
- Pooley, Oliver (2021). “The hole argument”. In: *The Routledge Companion to Philosophy of Physics*. Routledge, pp. 145–159. DOI: 10.4324/9781315623818.

- Russell, Bertrand (1901). “Recent work on the principles of mathematics”. In: *The International Monthly* 4, pp. 83–101.
- Sider, Theodore (2020). *The tools of metaphysics and the metaphysics of science*. Oxford University Press.
- Smart, J.J.C. (1961). “Free-will, praise and blame”. In: *Mind* 70.279, pp. 291–306. DOI: 10.1093/mind/lxx.279.291.
- Teitel, Trevor (2019). “Holes in spacetime: Some neglected essentials”. In: *Journal of Philosophy* 116, pp. 353–389. DOI: 10.5840/jphil2019116723.
- (2021). “What theoretical equivalence could not be”. In: *Philosophical Studies* 178.12, pp. 4119–4149. DOI: 10.1007/s11098-021-01639-8.
- Weatherall, JO (2018). “Regarding the ‘hole argument’”. In: *The British journal for the philosophy of science* 69, pp. 329–350. DOI: 10.1093/bjps/axw012.
- (2021). “Classical spacetime structure”. In: *The Routledge Companion to Philosophy of Physics*. Routledge, pp. 33–45.