

From statistical dependence to the space of possible superdeterministic theories*

Mordecai Waegell¹ and Kelvin J. McQueen^{1,2}

¹Institute for Quantum Studies, Chapman University

²Philosophy Department, Chapman University

September 30, 2025

Abstract

Bell's theorem demonstrates that any physical theory that is consistent with the predictions of quantum mechanics, and which satisfies some apparently innocuous assumptions, must violate the principle of *local causality*. It may therefore be possible to maintain local causality by rejecting one of these other assumptions instead. One possibility that has recently received significant attention involves rejecting the principle of *statistical independence* (SI). In this paper, we consider the *frequency interpretation* of SI, which states that $\rho(\lambda) \approx \rho(\lambda|Z)$, where $\rho(\lambda)$ is the relative frequency of an element of an ensemble being in the state λ , and Z is a label that separates the ensemble into apparently randomly selected sub-ensembles. SI is violated when the sub-ensemble frequency $\rho(\lambda|Z)$ fails to be representative of the ensemble frequency $\rho(\lambda)$. We argue that physical theories that systematically violate SI should all be understood as *superdeterministic*. This perspective on SI sheds light on a number of issues that are being debated in the superdeterminism literature, especially concerning its scope and philosophical consequences. Regarding scope, we argue that superdeterministic theories fall into three categories, deterministic theories with fine-tuned initial conditions, fluke theories, and nomic exclusion theories. We also argue that retrocausal and invariant set theories need not violate SI, which is contrary to how they are normally presented. Regarding philosophical implications, we argue that superdeterminism is incompatible with free will according to some prominent compatibilist accounts. We also argue that although superdeterminism is conspiratorial, it is not unscientific, but pre-scientific.

Keywords: Superdeterminism, statistical independence, measurement independence, Bell's theorem, locality, free will, retrocausality.

*This paper is forthcoming in *The European Journal for Philosophy of Science*.

Contents

1	Introduction	1
2	From statistical dependence to superdeterminism	2
2.1	What is statistical independence?	2
2.2	Three forms of superdeterminism	5
2.2.1	Determinism with fine-tuned initial conditions	5
2.2.2	Statistical flukes	7
2.2.3	Nomic exclusion	9
2.3	Do retrocausal and invariant set theories violate SI?	12
2.3.1	Retrocausal theories	12
2.3.2	Invariant set theory	14
3	Philosophical consequences of superdeterminism	16
3.1	Is superdeterminism compatible with free will?	16
3.2	Is superdeterminism conspiratorial?	19
3.3	Is superdeterminism unscientific?	21
4	Conclusion	22

1 Introduction

Bell’s theorem (Bell 1964, Bell 1971) shows that any physical theory that is consistent with the predictions of quantum mechanics, and which satisfies some apparently innocuous assumptions, must violate the principle of *local causality*.

Some argue that these other assumptions are *so* innocuous that they cannot coherently be denied, so that we may confidently assert that Bell’s theorem has *established* that physics is nonlocal (Maudlin (2014)). Others think that local causality is worth preserving, leading them to question these other assumptions, and to inquire into the feasibility of a local quantum theory that denies one of them. The apparently innocuous assumption that is the focus of this paper is the principle of *statistical independence* (or SI for short).¹

There has been much debate over how to best understand SI. We think SI is best understood as stating that *random sampling procedures yield representative samples*, and that viewing it this way helps to clarify its consequences. This is a familiar idea in science. It means that we can collect up entities of some type, so that experiments on those entities can potentially reveal general facts about *all* entities of that type. It is the basis of inductive generalization. It is no wonder that some have thought that denying SI would “undercut the whole scientific method” (Maudlin 2014). More recently, several authors have tried to make this objection more explicit (Baas and Le Bihan 2023, Allori 2024).

Others disagree, and believe that so-called *superdeterministic theories* can be rigorously developed. These are theories that deny SI (instead of local causality), and which aim to explain the outcomes of Bell experiments.² Some have also offered explicit responses to the above-mentioned criticisms of superdeterminism (Hossenfelder and Palmer 2020, Andreoletti and Vervoort 2022, Nikolaev and Vervoort 2023).

In light of these developments a number of debates have arisen about superdeterministic theories, what they do and do not require, whether they allow free will and whether they are conspiratorial. Our aim in this paper is to make clear what it really means for a theory to violate SI, i.e., to imply that random sampling procedures fail to give representative samples (especially in the context of Bell experiments). In doing so, we aim to contribute to all of the above mentioned issues.

In our view, Bell coined the term ‘superdeterministic’ to refer specifically to the idea that nature may not provide sufficient sources of randomness, such that random selection processes in experiments could never be relied upon to give representative samples. This is the exact scope of SI, so it seems natural to equate SI violation with superdeterminism as we do here. We think this passage from Bell (1990) makes this clear:

“An essential element in the reasoning here is that [the measurement settings] are free variables. One can envisage then theories in which there just are no free variables

¹Other options for preserving locality include denying the assumption that there is only one world (Vaidman 2016, Waegell and McQueen 2020) and denying that there is no retrocausality (see section 2.3.1 below).

²Attempts at developing superdeterministic theories can be found in t’Hooft (2016), Donadi and Hossenfelder (2020), Palmer (2022, 2024), and Ciepielewski, Okon, and Sudarsky (2023).

[...] In such ‘superdeterministic’ theories the apparent free will of experimenters, and any other apparent randomness, would be illusory. Perhaps such a theory could be both locally causal and in agreement with quantum mechanical predictions. However I do not expect to see a serious theory of this kind. I would expect a serious theory to permit ‘deterministic chaos’ or ‘pseudorandomness’, for complicated subsystems (e.g. computers) which would provide variables sufficiently free for the purpose at hand. But I do not have a theorem about that.”

Bell is particularly concerned here about getting a sufficient type of randomness out of deterministic theories so that they are not also superdeterministic. The frequency interpretation of SI addresses this concern, since given a sub-ensemble along with the ensemble from which it was drawn, one can evaluate SI, and it does not matter whether those ensembles were selected using a deterministic or indeterministic process. The success or failure of a random selection process to give a representative sample is independent of whether or not nature is deterministic.

In section 2 we build on existing work that aims to show that SI is not a probabilistic claim, but is instead a claim about whether or not our samples are representative of the populations we have sampled from (Chen 2021, Ciepielewski, Okon, and Sudarsky 2023, Allori 2024). But we do so in a way that enables us to challenge the widespread claim that superdeterminism is restricted to deterministic theories with fine-tuned initial conditions. Section 2.2.1 explains theories of this sort and responds to arguments that superdeterminism must take this form. Section 2.2.2 explains a quite different class of superdeterministic theories based on statistical flukes. Section 2.2.3 explains a third category of superdeterminism, which we call “nomic exclusion” theories.

Section 2.3 then takes two different types of theories that are often said to violate SI (retrocausal theories and invariant set theory), and argues that this need not be the case, and that neither type of theory is necessarily superdeterministic.

Section 3 then moves to philosophical consequences of superdeterminism. In section 3.1 we challenge the widely held claim that superdeterminism has no bearing on free will. In particular, we argue that some prominent compatibilist accounts of free will entail that superdeterminism restricts our free will. Section 3.2 argues that superdeterministic theories all have a conspiratorial element to them, and we respond to arguments to the contrary (Andreoletti and Vervoort 2022). Finally, section 3.3 addresses the scientific status of superdeterminism. We argue that although superdeterminism does not necessarily undermine the scientific method, that does not make it scientific. We think it is best understood to be in a pre-scientific state.

2 From statistical dependence to superdeterminism

2.1 What is statistical independence?

The assumption of statistical independence (SI) was first made explicit by Clauser and Horne (1974, note 13). Bell (1977) explicitly acknowledges the need for it, but describes it as “a point of leverage for ‘free willed experimenters’”. As a consequence, SI has sometimes been called *the*

Free Will assumption or the *the Freedom of Choice Assumption* (Myrvold, Genovese, and Shimony 2020). Recent research has almost unanimously rejected Bell’s characterization, arguing that free will has nothing to do with it, a point we return to in section 3.1.

SI is now standardly defined as $\rho(\lambda|X,Y) = \rho(\lambda)$, where X and Y are measurement settings and λ are what determine the measurable properties of the measured systems. However, there have been different interpretations of ρ : is it a probability or is it a relative frequency? And if it is a probability, is it an epistemic or an ontic probability?

The probabilistic interpretation has a long history. For example, as Chen (2021) explains, many sought to maintain *both* local causality and SI, by rejecting implicit assumptions about classical probability theory that Bell may have made in his proof. Chen responds by proving Bell’s theorem without using probability theory. In particular, Chen formulates each assumption of Bell’s theorem without using probabilities, and in the case of SI, he adopts the frequency interpretation. The probabilistic interpretation has also been criticized by Ciepielewski, Okon, and Sudarsky (2023, p444), who instead defend the frequency interpretation. The frequency interpretation has also recently been adopted and defended by Allori (2024)).³ We find that this interpretation of ρ provides a fresh perspective on many topics discussed in the context of superdeterminism and our goal in this paper is to clarify the interpretation and its consequences for these topics.

On the relative frequency interpretation, SI is easy to interpret, and applies well beyond quantum mechanics. Simplified, it states that $\rho(\lambda) \approx \rho(\lambda|Z)$, where $\rho(\lambda)$ is the relative frequency that an element of an ensemble is in the state λ , and Z is a label that separates the ensemble into apparently randomly selected sub-ensembles.

For example, in an ensemble of apples, λ might be ‘red apple’, and Z might be ‘left box’. SI then says that if the sub-ensembles are sufficiently large, and selected in an apparently random way, then the relative frequency of λ in every sub-ensemble should be approximately the same as its relative frequency in the entire ensemble. So, if half the apples are red, then approximately half the apples in the left box are red. This is really just another way of saying that the random sampling procedure marked by Z generates a *representative sample* of a larger ensemble. Consequently, a theory which systematically violates SI, that is, a superdeterministic theory, must entail that the random sampling procedure fails to yield representative samples.

Random sampling procedures are a familiar practice in science. They begin by defining the total ensemble or population of interest, which we want to learn general facts about. Typically, the population of interest is too large for us to examine every member. And so we instead only examine a sample of the population. If the sample is representative of the population, then we can be confident that what we discover about our sample represents a discovery about the total population. Acquiring a *random* sample is a way to help ensure our sample is representative of the population. As Chen (2021) and Allori (2024) emphasize, random sampling is what allows scientists to use induction.

How do we ensure that our samples are randomly selected from the population of interest?

³An early defense of the frequency approach is given by Maudlin in an illuminating online debate with Tim Palmer who, as we shall see in section 2.3.2, instead adopts the ontic probability approach. See youtube.com/watch?v=883R3JIZHXE.

Consider a simple yet notorious example. In 1936, the magazine *Literary Digest* obtained a sizable sample of polling data of the voting population of America: 2.4 million voters. In this sample, 57% would vote for the Republican, Alf Landon, so the magazine predicted a loss for the Democrat, Franklin Roosevelt. Yet Roosevelt won 62% to 37%. This is a case in which random sampling *failed to take place for the intended population*. The sampling method that was used meant that effectively everyone in the sample was wealthy. Since wealth is a relevant factor for voting preference, random sampling over the US population did not take place. Instead, random sampling only happened (if at all) over the *wealthy* U.S. population. This is an important distinction. For as we understand superdeterminism, representative samples are not obtained *even when random sampling over the relevant population is apparently successful*, e.g., even when the sample includes a proportional representation of wealthy and non-wealthy people (and all other observable properties). In other words, if superdeterminism is true, there isn't anything we can do to obtain a representative sample (for the relevant domain of interest).

SI is relevant to Bell experiments because to conclude that nature is nonlocal, we assume that the choice of measurement settings is independent of the state of the system to be measured. If our measurement settings are the angles at which we position our polarizers, and our measurement outcomes are whether or not entangled photons are absorbed by those polarizers, then it is essential that the settings are chosen independently of the properties of the photons used in the experiment. This independence prevents any systematic bias that could lead to unrepresentative samples—where, for instance, photons sent to a polarizer at one angle are fundamentally different from those sent to a polarizer at another angle. To achieve this, settings must be selected randomly. Various methods are employed to ensure randomness in Bell experiments: some use pseudo-random number generators, which rely on algorithms to produce sequences of bits that are statistically independent; others employ quantum processes involving inherently unpredictable outcomes according to quantum theory. More innovative approaches include using cosmic photons, such as signals from distant stars or quasars, where the vast distances and time intervals make any causal link between these photons and the experiment highly implausible. These strategies are designed to increase our confidence that the settings are chosen randomly and independently of the photon state, and thus that SI is obeyed for these Bell experiments.

When we perform the above-mentioned experiments on certain sorts of entangled photon pairs, we see certain very specific results. For example, if the two polarizers point in the same direction (modulo GR corrections), then the measurement outcomes on the two photons will be the same: either both will be absorbed or both will be transmitted (Chen 2021). When Bell's theorem is proved in the context of this type of experiment, it must be assumed that these results are indicative of all photons, and not just the ones that happened to be measured in this way. Superdeterministic theories deny this assumption.

It is widely believed that superdeterminism must be deterministic and must involve fine-tuned or atypical initial conditions. For example, Baas and Le Bihan (2023) assert that “superdeterminism is the conjunction of determinism and the atypicality of cosmological initial conditions”. Meanwhile, Allori (2024, p156) asserts that “superdeterminism can succeed only [by] resorting to

fine-tuned initial conditions”. (Allori’s claim is based on a formal result by Dürr and Teufel (2009), which we discuss in the next section.)

But this is an unnecessary restriction that can lead one astray when making general claims about superdeterminism, as we shall see. Superdeterminism was introduced by Bell as a loophole to Bell’s theorem, one that allows us to maintain locality by rejecting SI. Fine-tuned initial conditions may be one way of reaching this result, but it is not the only way.

We have argued that the key defining feature of superdeterminism is the violation of SI.⁴ We think there are two ways other than fine-tuned initial conditions that can yield this result. We therefore propose the following:

A theory is superdeterministic if and only if it entails that random sampling procedures systematically fail to yield representative samples, in virtue of either

- 1) Fine-tuned initial conditions in a deterministic theory; or
- 2) Statistical flukes (in a deterministic or indeterministic theory); or
- 3) Nomic exclusion (in a deterministic or indeterministic theory).

In the next section (2.2), we consider each of these in turn. Then in the following section 2.3 we consider two cases that are difficult to categorize, retrocausality and invariant set theory.

2.2 Three forms of superdeterminism

2.2.1 Determinism with fine-tuned initial conditions

Superdeterminism is often taken to entail fine-tuned or atypical initial conditions in a deterministic universe. Take the fact that for the relevant entangled photon pairs, we always see that if the two polarizers point in the same direction, then either both photons will be absorbed or both will be transmitted. Perhaps only some photons behave this way, while most don’t, it’s just that the ones that don’t never make their way into this type of experimental set up, due to the atypical initial conditions of the universe.

Is it inevitable that superdeterminism be a deterministic, local theory with fine-tuned atypical initial conditions? While many have taken these to be defining features of superdeterminism, we think none of them are necessary and in the next two sections we present counterexamples. But first we should consider a result in Dürr and Teufel (2009), which Allori (2024) has described as “a formal result establishing that SI holds for typical initial conditions, so that superdeterminism can succeed only [by] resorting to fine-tuned initial conditions”.

Dürr and Teufel (2009) demonstrate a result for deterministic frameworks, using the example of a Galton board—a device where balls bounce off pegs and eventually settle into boxes at the bottom, forming a distribution that matches probabilistic predictions. In their analysis, they trace the randomness of the balls’ paths back to the initial conditions of the balls entering the board and

⁴Several authors have put forth their own proposed definitions of superdeterminism e.g. Wiseman and Cavalcanti (2017), Waegell and McQueen (2020), Sen and Valentini (2020a), Wharton and Argaman (2020), Adlam et al. (2024), which are each distinct in some way.

further to the cosmological initial conditions of the universe. By introducing a typicality measure for these cosmological initial conditions, they show that typical initial conditions will lead to a ball bouncing left or right off the pegs it encounters a roughly equal number of times, whereas atypical initial conditions could result in a ball bouncing left off of every peg in the board. Dürr and Teufel do not explicitly connect this result to SI. But perhaps Allori’s thought is that if we were to see the balls always bouncing left, then we would have a sample that fails to represent typical initial conditions, in virtue of atypical fine-tuned initial conditions.

However, even if Allori’s interpretation of Dürr and Teufel’s result were correct, it still falls short of showing that superdeterminism (the denial of SI) entails atypical or fine-tuned initial conditions. First, the result assumes determinism, but in section 2.2.2 we show how indeterminism without fine-tuned initial conditions can violate SI. Second, even assuming determinism, it assumes there are no special systems (“goblins”) whose behavior leads to SI violations, without atypical initial conditions. This case is explained in section 2.2.3.

In the remainder of this section, we consider a recent specific superdeterministic model, based on fine-tuned initial conditions (Ciepielewski, Okon, and Sudarsky 2023). Part of the interest of this model is that it postulates atypical initial conditions that are not very complicated. It therefore offers a response to a common criticism of superdeterminism that it requires overly complex initial conditions (Chen 2021). Following Chen (2021), we will refer to it as Leibnizian Quantum Mechanics (LQM), due to some parallels to Leibniz’s Monadology (Leibniz 1714). LQM is a superdeterministic model that replicates quantum predictions using a unique structure where, at each point in physical space, there exists an internal space (a copy or simulation of a universe). In LQM, the evolution of the universe in physical space is determined by the dynamics within these internal spaces, which each follow Bohmian mechanics. Essentially, LQM replaces the single universal wave function in Bohmian mechanics with a continuous infinity of wave functions, each belonging to an internal space and evolving independently within that space following the Schrödinger equation. There are no particles moving in physical space, and instead the configuration of mass density at a given location in physical space is given by the internal state at the corresponding location. The initial conditions are atypical in that they are identical for every internal space and there is no reason to expect this, given that they are all independent. This model shows that the atypical initial conditions of a superdeterministic theory need not be overly complex, although as Chen (2021) notes, an excessive complexity lies instead in the Leibnizian ontology.

How exactly is this model superdeterministic? According to the authors, “To see that SI is indeed violated once homogeneous conditions are provided, we note that the state λ of the particles involves the specification of both Φ (the universal wavefunction) and Z (the Bohmian position variables) in the region where the particles are created. But the homogeneity of initial conditions implies that such Φ and Z determine such fields everywhere, including in the regions where a and b (settings) are located. We conclude that in this case, λ , a , and b cannot be independent” (Ciepielewski, Okon, and Sudarsky 2023, p454). Bohmian mechanics, as typically understood, does not violate SI. This is because the wavefunction of the Bell state (λ) and the experimental settings (a and b) are (effectively) separable from one another and can be independently varied

(Esfeld 2015, Allori 2024). LQM, on the other hand, apparently does violate SI, despite the similar ontology. This is because the λ that is associated with a particle corresponds to the Φ and Z in the region of the particle, and these correspond to an *entire internal Bohmian universe*. Given homogeneous initial conditions, every point in physical space has the same internal state Φ and hidden variable Z , so the λ of any given particle encodes information about all settings, no matter where in physical space they are chosen. Thus, given the atypical homogeneity condition, neither the settings nor λ can be varied, no matter where they are in physical space, so they remain perfectly correlated. For a typical inhomogeneous initial condition, no such correlations would exist, since the wavefunction at one location generally encodes no information about any other location - and it seems physical space is generally an abstract mess, which contains nothing resembling a Bell experiment.

However, the authors adopt the frequency interpretation of SI, in which $\rho(\lambda)$ represents “the actual distribution of states over the measured ensemble” (p444). But in the explanation of the violation of SI quoted above, λ is understood in terms of Φ and Z , i.e., entire internal Bohmian universes, which given homogeneous initial conditions, cannot be different for different particles. So there is a problem here in explaining how LQM violates SI under their preferred frequency interpretation. Part of the difficulty here is the very exotic ontology on which the theory is based. After all, what does the ontology look like given inhomogeneous initial conditions? Getting clear on exactly how the model violates SI in the preferred sense is crucial, as a key claim is that violations of SI are restricted to quantum experiments, and do not arise, for example, in randomized clinical trials.⁵ This point is crucial to whether superdeterminism can be considered a truly scientific theory (discussed in section 3.3). We do not think this is impossible to show, but think more work needs to be done to make this clear: given this ontology, what exactly does a distribution of λ s and corresponding settings a and b look like for trials of a Bell experiment, so that we can see an explicit correlation between them? Either way, we agree that LQM provides a useful framework to advance the debate over superdeterministic violations of statistical independence.

2.2.2 Statistical flukes

This section considers a class of superdeterministic theories that do not involve fine-tuned initial conditions, but instead violate SI through statistical flukes. There are both deterministic and indeterministic examples.

To understand how this is possible, we first must be clear on what is meant by “theory”. In the previous section, we discussed LQM. However, LQM given inhomogeneous initial conditions does not violate SI. What violates SI is LQM plus an additional postulate: homogeneous initial conditions. So strictly speaking, the “theory” that violates SI is really a physical theory (in the ordinary sense, i.e., dynamics without boundary conditions) plus a postulate. The same idea will apply here in this section, except instead of an additional postulate about initial conditions, we have additional postulates about flukes. A superdeterministic theory based on flukes is therefore

⁵See Ciepielewski, Okon, and Sudarsky (2023, p460). See also Hossenfelder (2020, sec. 6), who appeals to LQM as a solution to this problem.

to be understood as a conjunction of a physical theory plus a fluke postulate. A fluke postulate does not concern an initial boundary condition on the events in spacetime, but the existence of an improbable distribution of events throughout spacetime.

To begin with, we show how a fluke postulate could violate SI (given the frequency interpretation). We will then move to superdeterministic theories. A simple indeterministic example involves the familiar case of flipping a fair coin. If we flip one million coins, it is overwhelmingly probable that the number of heads and tails outcomes is roughly equal, which gives a representative sample of the probability distribution we associate with flipping fair coins. But it is also possible to get all heads, or all tails, or other highly improbable fluke cases that are not representative of the underlying 50/50 probability distribution that was used to randomly generate the sample. If one flips a large number of fair coins and obtains such a highly improbable result, then random sampling has failed to yield a representative sample.

This is an important example, because we can see the violation of SI in two different ways. First, in the ensemble sense, when we look at typical data from a billion or more actual coin flips, and see that our million heads really do fail to represent the roughly 50/50 distribution in the larger ensemble, and second, without reference to a larger ensemble, we see that the million heads fail to represent the underlying 50/50 probability distribution used to generate the individual outcomes. In the second case, we have no empirical reference that tells us the distribution should be 50/50, but if we are proposing a particular physical theory with a particular distribution (e.g., 50/50 for fair coin flips), we can see if the ensemble is representative of the proposed distribution. It is also possible that all coins ever flipped will be heads (a ubiquitous fluke), meaning there is no larger ensemble that the million heads fail to represent, so for this case, we can only see the violation of SI in the second way.

It could likewise be that the fact we always see entanglement correlations obeyed in a quantum experiment is a statistical fluke in an indeterministic theory where the correlations are only obeyed with some probability. Consider such a theory where every time we prepare an entangled state and measure it, the quantum correlations are obeyed for half of all cases, and violated for the other half. Then seeing the entanglement correlations obeyed a million times in a row is a fluke in exactly the same way as seeing the million heads. That is, our experimental data were not representative of the true underlying theory (i.e., SI is violated), and have misled us into deriving quantum theory.

As a final indeterministic example, take any deterministic theory that is superdeterministic because of a fine-tuned initial condition, but change it so that the initial condition is selected at random from a distribution for which this fine-tuned initial condition is atypical. In that case, the fact that we see entanglement correlations obeyed in Bell experiments would be a statistical fluke, rather than a fine-tuning, even though the initial conditions and the resulting universe are the same. Note that for some indeterministic theories, we can obtain a deterministic counterpart theory by reinterpreting the indeterministic events in the former as boundary conditions in the latter. In that case, a fluke in the indeterministic theory becomes a fine-tuned boundary condition in the deterministic counterpart.

Deterministic branching theories also produce statistical flukes. A simple example is based on

the Everett interpretation (as commonly understood e.g. as in Wallace (2012)), which implies that we have arrived at quantum mechanics in part because we live in a branch whose measurement results confirm the Born rule. There are many other branches in which observers see different results, and so arrive at a different (false) physical theory, or no theory at all (Hemmo and Pitowsky 2007). To simplify, if the coins in the above example were quantum coins with 50/50 Born rule probability to be heads or tails, then there is a world with all heads, and another with all tails, even though there are far more worlds with approximately equal numbers. If an observer finds themselves in a world with all heads, then SI is violated for the same reasons as above, for that observer. The million heads is a fluke because it fails to represent the 50/50 Born rule distribution that was used to generate all of the 2^{10^6} worlds. It would also fail to represent a billion quantum coin flips in the vast majority of worlds, where the number of heads and tails is nearly equal. And there is also one world where the fluke is ubiquitous, and all quantum coins ever flipped come up heads. In this world, there is no larger ensemble with roughly 50/50 outcomes that we can say this million coins fail to represent. All we can say is they fail to represent the underlying Born rule probability in the physical theory. In this theory, observers in the non-fluke branches are likely to derive quantum theory, while observers in the fluke branches will derive some other theory (e.g., a theory where all quantum coins come up heads).

Finally, there could be a deterministic branching version of the above theory where entanglement correlation rules are not generally obeyed, but can be obeyed for atypical branches. In this theory, entanglement correlations are obeyed in the atypical fluke branches, and so observers in these branches are likely to derive quantum theory, while observers in the non-fluke branches will derive some other theory (e.g., a theory where entanglement correlations are not obeyed). Note that atypical fluke branches will occur even for typical initial conditions, so these cases are unrelated to fine-tuning.

These approaches may seem so implausible that they should be left out of the definition of superdeterminism, but we see no strong reason to think that these theories are any less plausible than deterministic theories with fine-tuned initial conditions. Even Ciepielewski, Okon, and Sudarsky (2023) say of their own model that it is not a serious competitor to more standard interpretations of quantum mechanics. And Chen (2021) concludes that it falls short of being an empirically adequate theory that is overall simpler and more attractive than well-known non-local interpretations. We will return to the issue of the tenability of superdeterministic theories in section 3. For now, we turn to our final class of superdeterministic theories.

2.2.3 Nomic exclusion

In Maudlin’s brief discussion of superdeterminism (Maudlin 2014), he mentions an alternative to “massive coincidence” superdeterminism, which all of the above-discussed cases could be seen as examples of. The alternative instead appeals to our “being manipulated”: somehow we are led to choose the settings we choose, or the random number generators we offload this task to are rigged, or some such. Such hypotheses need not involve fine-tuned initial conditions, nor determinism, nor flukes, as we shall show. They may appear highly conspiratorial. But they may also be no

worse off in this regard than massive coincidence approaches, and some existing models appear to fall into this category.

Our third class of superdeterminism violates SI by making certain combinations of measurement settings and ontic states nomologically impossible. We call these “nomic exclusion models”. They allow for the observed data to be explained by a local hidden variable model. Any complete assignment of local hidden variables to the observables of an entangled state will violate quantum predictions for some measurement settings. To prevent this, these models make it physically impossible to measure those settings, so that quantum predictions are obeyed, but no random selection process will ever obtain a representative sample of the true local hidden variable distribution. Some models in this class simply assign no values to observables that are impossible to measure.

Before we consider existing superdeterministic models that take this form, we consider some fanciful toy models, that abstract away the complex details, and so help us to see exactly how nomic exclusion models violate SI. Our toy models appeal to *goblins*, which are somewhat akin to Maxwell’s Demon, who actively manipulate our measurement settings, and/or the hidden variables λ that determine the outcomes, in such a way that they are dependent - even if there is no empirical evidence of this manipulation. The goblins are effectively placeholders for the mechanisms of more sophisticated nomic exclusion models.

The goblins manipulate nature in much the same way that a stage magician deceives their audience. For example, a stage magician might ensure you choose a particular card from a deck by some trick, even though you think you are free to choose any card. The key difference is that it is possible with extremely careful observation to see how a stage magician’s trick is done, but there is no empirical evidence of the goblin’s tricks. A stage magician’s tricks can succeed with no fine-tuned initial conditions of the universe, nor statistical flukes, and the goblin’s tricks are no different in this regard.

The goblin theories obey local causality and do not involve retrocausality, because the goblins perform their nefarious deeds at subluminal speeds moving forward in time. In general, the goblins must begin their activities at a past event where they can act as a local common cause for the entire experiment, including the choices of measurement settings. In this way, the goblins use local hidden variables to produce correlations consistent with the quantum predictions for entangled states, including Bell inequality violations.

Nomic exclusion theories can themselves be categorized into three different forms, depending on the causal order in which the SI-violating correlations are set up. This may not be obvious when looking at existing sophisticated models, but our simple Goblin models help make these structural differences apparent. The Goblin models involve hidden variables (λ), measurement settings (X) of multiple experimenters, and goblins (G). They differ in terms of what causes what, where causal relations between λ , X , and G are represented below by arrows. We consider three different roles for the goblins, corresponding to three different types of superdeterministic theories. In all three cases, SI is violated because, for any experiment, not all measurement settings are assigned outcomes consistent with quantum predictions, but the goblins ensure no inconsistent settings can be chosen, thereby preventing the measurement of representative samples that would

reveal the discrepancy.

In what follows we refer to outcomes that correspond to our empirical observations (i.e., outcomes consistent with quantum predictions), as *empirically plausible* outcomes. Our goblin theories must assign empirically *implausible* outcomes to *some* measurement settings, because any set of local hidden variables for all measurement settings must violate some of the entanglement correlations.

Theory 1: $\lambda \longrightarrow G \longrightarrow X$.

Goblins examine the system's hidden variables to determine which settings have empirically plausible outcomes and which do not. They then tamper with the experimenter's brains (or their random number generators, etc.) to ensure that only settings with empirically plausible outcomes are possible. This is a clear case where measurement settings depend on (what determines) the measurement outcomes.

Theory 2: $\lambda \longleftarrow G \longrightarrow X$.

Goblins choose the system's hidden variables and create empirically plausible outcomes for some settings but not for others. They then tamper with the experimenter's brains (etc.) to ensure that only settings for which they created empirically plausible outcomes are possible. Here G is a local common cause of X and λ .

Theory 3: $X \longrightarrow G \longrightarrow \lambda$.

Goblins first examine physical systems that fully determine which future settings the experimenters choose. The goblins then create an empirically plausible outcome for only that joint setting. This can be generalized to the case where the future settings are restricted, but not fully fixed, in such a way that a local hidden variable theory can assign empirically plausible outcomes to all of the allowed settings.⁶

The goblin theories are empirically consistent because the empirically implausible outcomes are systematically hidden from the experimenters. This means the experimenters never get representative samples of the outcomes assigned to all settings, which violates SI. The goblin theories also help to illustrate that neither determinism with fine-tuned initial conditions, nor flukes, are necessary for the violation of SI. If the universe - goblins included - is a fully deterministic system, we see no reason to think that the initial conditions must be fine-tuned or atypical for the goblins to act this way. Besides, our goblins need not be deterministic, and might be acting this way due to their own libertarian free will.

As with superdeterminism based on statistical flukes, superdeterminism based on goblins (or some less fanciful but functionally equivalent mechanism) may seem too implausible, and so not deserving of the label superdeterminism. But again we appeal to the *tu quoque* response: it is far from obvious that fine-tuned initial conditions are any better off. In fact it could be noted that some authors have taken seriously the possibility that we are in a simulation (Bostrom 2003, Chalmers 2022); our simulators may act like goblins, to prevent us from discovering the underlying code of

⁶It has been claimed that only a retrocausal theory can depend on the settings in this way (Wharton and Argaman (2020) and Hance (2024)). However, it is in principle possible for goblins to use information from the past to predict what settings will be chosen with certainty, so that the Goblins can then choose hidden variables that give empirically plausible outcomes for these settings.

the simulation.

We now move on from the simple goblin toy models to more sophisticated models that have been proposed in the literature, which fall under the nomic exclusion category. The models of Donadi and Hossenfelder (2020) and Hance and Hossenfelder (2022) are consistent with goblin theory 2 ($\lambda \leftarrow G \rightarrow X$) and goblin theory 3 ($X \rightarrow G \rightarrow \lambda$). For in these models, the physical state λ at the source depends explicitly on the future setting choice X , and this state does not specify outcomes for any other measurement settings. Therefore, no other choices of the measurement settings are physically possible. They do not specify how this dependence comes about in their model. In theory 2, using the goblin visualization for simplicity, a goblin chooses and fixes the future setting X well in advance, and then prepares the ontic state λ so that entanglement correlations will be obeyed for setting X . Outcomes for other measurement settings are not even defined for this λ . The goblin still produces the standard quantum long-run statistics for random choices of measurement settings and their outcomes, but SI is clearly violated because the cases that are measured by each different setting X have distinctly tailored λ (i.e., $\rho(\lambda|X) \neq \rho(\lambda|X')$).

The counterfactually restricted theory of Hance (2024), which is closely related to Hance, Hossenfelder, and Palmer (2022), is a nomic exclusion model that is consistent with goblin theory 1, since observables with empirically implausible outcomes are nomologically impossible to measure, while it is possible to measure any observables with plausible outcomes. We discuss the (2022) model in more detail in section 2.3.2.

Finally, Ciepielewski, Okon, and Sudarsky (2023) consider an alternative version of LQM, where the superdeterminism does not arise from initial conditions or flukes, but instead from the laws. In particular, they start from generic initial conditions and then show that, from the dynamics alone, one can arrive at the condition of homogeneity that correlates settings and λ . Another way to say this is that the homogeneous state is an attractor of the theory, which it will settle into for a wide variety of typical initial conditions. This can be seen as a version of Goblin theory 2, where the Goblins (here, the laws) act over time to create superdeterministic correlations.

2.3 Do retrocausal and invariant set theories violate SI?

2.3.1 Retrocausal theories

It is often said that there are two quite distinct ways of maintaining locality by violating SI, one is superdeterminism, the other appeals to *retrocausality*. Since we have defined all systematic violations of SI to imply superdeterminism, this would mean that retrocausality is also a type of superdeterminism.⁷ However, in what follows, we argue that retrocausality, properly understood, does not violate SI, and it is for this reason that retrocausality is not superdeterministic.

We will argue that there are two different ways of interpreting retrocausal models, in terms of *causal order* or *temporal order* and that while SI does appear violated in the more standard temporal order interpretation, the causal order interpretation is preferable, and need not violate

⁷Myrvold, Genovese, and Shimony (2020) and Ciepielewski, Okon, and Sudarsky (2023) are examples of the standard view, according to which retrocausality violates SI without being superdeterministic. Nikolaev and Vervoort (2023) is an exception, as they treat retrocausality as superdeterministic.

SI. The λ s in $\rho(\lambda) = \rho(\lambda|Z)$ are supposed to represent the relevant ontic states provided by the physical theory. A key premise in our argument will be that the ontic states relevant to SI will generally differ between the causal and temporal order interpretations. We should also stress that the ontic states λ in retrocausal theories do not obey Bell's notion of local causality, which are predicated on causal influences obeying temporal order and never connecting spacelike separated events.

As a simple example of a retrocausal toy model, consider the Zig-Zag model of Costa de Beauregard (1976).⁸ In this model of a Bell experiment, the causal chain begins at the source, goes to Alice's measurement, goes back (in time) to the source, and then finally goes to Bob's measurement, with none of these links directly connecting space-like separated events. This model explains the results of the Bell experiment using local retrocausality, and the entanglement correlation is produced because Alice's result is sent back to the source where the other qubit is then prepared in the properly correlated state before being sent to Bob.

What follows is a formalization of this model using standard Born rule probabilities, but for local ontic states. Given the prepared Bell state $|\Psi\rangle_{12} = \frac{1}{\sqrt{2}}(|0\rangle_1|1\rangle_2 - |1\rangle_1|0\rangle_2)$, the reduced density matrix of the first qubit is the maximally mixed state $\hat{\rho}_1$. This is defined to be the local ontic state of Alice's qubit in this model, i.e., $\hat{\rho}_1 = \lambda_1$. Alice chooses a measurement setting X and measures her qubit, obtaining an outcome with standard Born rule probability given $\hat{\rho}_1$. Alice's outcome state $|k\rangle_1$ is sent retrocausally back to the source event, and the other qubit is then prepared in the properly correlated state obtained by projecting the Bell state onto Alice's outcome and renormalizing, $|\phi\rangle_2 = \sqrt{2}\langle k|_1|\Psi\rangle_{12}$. That qubit is then sent to Bob, and this quantum state is defined to be the local ontic state $|\phi\rangle_2 = \lambda_2$ of that qubit. Bob chooses a measurement setting Y and measures his qubit, obtaining an outcome with standard Born rule probability given $|\phi\rangle_2$. Bob's outcome obeys the entanglement correlations because of how $|\phi\rangle_2$ is defined from the Bell state using Alice's outcome.

We now show how SI is violated by our retrocausal Zig-Zag under the standard temporal order interpretation. In this picture, both λ_1 and λ_2 exist at the source, so the complete ontic state at the source is $\lambda = \lambda_1 \cup \lambda_2$. The values of λ_2 are perfectly correlated with Alice's measurement setting X , because they are created using the outcome of Alice's measurement. As a result, the equation $\rho(\lambda_2|X) = \rho(\lambda_2)$ is violated, and thus by extension the SI equation $\rho(\lambda|X) = \rho(\lambda)$ is also violated. In other words, if Alice randomly separates her ensemble into two sub-ensembles, where all elements in one bin will be measured by one setting, and all elements in the other by another setting, the λ_2 s in each sub-ensembles would all match (predetermine) her setting, and the two sub-ensembles would not be representative of the overall distribution of λ_2 values in the ensemble.

We now show how SI is *not* violated under the causal order interpretation. At the first step in the causal order of this model, the state λ_1 has been prepared, but λ_2 does not yet exist, because Alice has not yet chosen her setting or made her measurement, so the complete ontic state at this

⁸For other retrocausal models, see Sutherland (1983), Sutherland (2008), Price and Wharton (2015), Wharton (2018), and Wharton and Argaman (2020).

step is $\lambda = \lambda_1$. According to this model, when Alice determines her choice of setting, she always gets representative samples of λ_1 for any setting, because $\rho(\lambda_1|X) = \rho(\lambda_1) = 1$, and it follows that the SI expression $\rho(\lambda|X) = \rho(\lambda)$ is obeyed. After Alice performs her measurement with setting X , and λ_2 is created using the outcome, $\rho(\lambda_2|X) = \rho(\lambda_2)$ is still violated as before, but this is just the expected correlation between a setting and an outcome *after* a measurement has been made, so does not actually indicate a violation of SI. In temporal order, λ_2 exists before X is chosen, but in the Zig-Zag causal order, λ_2 only exists after Alice’s measurement has been performed. A simple way to see why violation of $\rho(\lambda_2|X) = \rho(\lambda_2)$ does not constitute an SI violation is that X is no longer a random selection after the measurement. By using selection X , one knows the λ_2 in each sub-ensemble will be eigenstates of a particular setting observable.

In the final step, Bob chooses his setting and performs his measurement. Even though λ_2 may take on different values from run to run, it is still the case that $\rho(\lambda_2|Y) = \rho(\lambda_2)$ because nothing depends retrocausally on Bob’s randomly selected measurement setting Y . As a result, Bob succeeds in obtaining a representative sample of the overall distribution of $\lambda = \lambda_1 \cup \lambda_2$, and thus SI is also obeyed for Bob’s measurement, i.e., $\rho(\lambda|Y) = \rho(\lambda)$.

Thus, we have shown that when SI is evaluated following the causal order of a retrocausal theory, it is not necessary violated (although a retrocausal theory could certainly be superdeterministic for other reasons, such as a fluke postulate). We conclude that because SI depends explicitly on the ontic states λ provided by a theory, it makes more sense to evaluate SI in the causal order in which λ evolves within that theory. This applies to all physical theories, not just this Zig-Zag model. We think that if we apply it to any theory, retrocausal or not, and find that SI is violated, it is correct to conclude that the theory is superdeterministic.

Proponents of retrocausal models have argued that generalized relativistic locality can be preserved, even allowing some causal influences to move backward in time between lightlike or time-like separated events, provided none ever directly connect spacelike separated events. This allows indirect causal chains to connect spacelike separated events, as in this example. The Zig-Zag model above obeys this generalized locality principle, as well as SI, so a reformulated version of Bell’s theorem using this locality principle would need a separate assumption to rule out retrocausal models. This is analogous to how Bell did not initially acknowledge SI as a necessary assumption, until it became clear that violating SI allows for local hidden variable theories.⁹

2.3.2 Invariant set theory

The “invariant set theory” model is defended in Hance, Hossenfelder, and Palmer (2022) and in Palmer (2024). The model is intended to be a local, superdeterministic interpretation of quantum mechanics. However, while we agree that the model presented in (2022) is local and superdeterministic, we find that the modifications proposed in (2024) result in a model that is neither local nor superdeterministic.

In the (2022) paper, the authors make clear that the existence of a specific invariant set λ makes

⁹This is also analogous to the eventual acknowledgment that a “one world” assumption is necessary too, see footnote 1.

it nomologically impossible for Alice and Bob to choose certain pairs of settings, which clearly leads to a violation of SI (see goblin theory 1 in section 2.2.3). The impossibility of those settings allows for an entirely local hidden variable model to explain the empirical data in a Bell experiment. In this version, the invariant set is determined before Alice and Bob make their choices.

In the more recent paper, Palmer (2024) insists that Alice and Bob are free to choose any setting, and that their choices are part of the all-at-once (across all of space and time) selection of a compatible invariant set, so their choices are *ontologically prior* to the selection of the invariant set - i.e., the choices are made ‘before’ the invariant set is determined. Palmer thus adopts an “all-at-once” model (Adlam 2022), where no “initial” condition or state is privileged and gives rise to later states. All states of the universe instead come together at once. This process involves synthesizing information from many spacelike separated events, and then distributing correlated outcomes to spacelike separated events, and is thus highly nonlocal.

In Palmer’s model, given choices X and Y , certain values of λ cannot be selected by nature. In other words, the order of ontological priority has been reversed, and the free choices X and Y are determined ‘before’ the invariant set λ .¹⁰

Palmer’s argument that his theory violates SI assumes that ρ is an ontic probability. Since his theory is deterministic, the ontic probabilities are all zero or one. His argument for the violation of SI can be concisely reconstructed as follows (where the measurement settings take binary values 0 and 1, and $Y' = 1 - Y$):

- (1) $\rho(\lambda|X, Y) = 1$. [From the theory’s determinism.]
- (2) $\rho(\lambda|X, Y') = 0$. [From the theory’s “counterfactual indefiniteness”.]
- (3) If SI then $\rho(\lambda|X, Y) = \rho(\lambda|X, Y')$.
- (4) So, SI is false.

Premise (3) holds because SI requires $\rho(\lambda) = \rho(\lambda|Z)$ for any apparently random selection Z .

Premises (1) and (2) come from invariant set theory. Alice and Bob measured X and Y which are the settings corresponding to ontic state λ . Had Bob measured Y' instead, the ontic state λ would have been different, because it is nomologically impossible to measure Y' given the original λ . Thus, if Alice and Bob are truly free to measure any settings, then a compatible value of λ must be determined nonlocally using both of their spacelike separated settings, such that the entanglement correlations are obeyed by the eigenvalues in λ . The eigenvalues of the measured settings are then nonlocally distributed to the spacelike separated measurement outcome events. Essentially, by selecting their settings X and Y , Alice and Bob obtain a correlated value of λ as an outcome.

Let us now explain how the all-at-once selection process is supposed to work in the model. First, to simplify, imagine that free settings X and Y are the only measurements ever performed in the universe. Then according to Palmer (2024), reality takes a certain form λ , which determines eigenvalues for these measurement settings that conform to quantum mechanics. This gives

¹⁰For more on “all-at-once” freedom, see Waegell, McQueen, and Adlam (2023, Sec. 5.2).

premise (1). This particular λ will not (according to this theory) give eigenvalues for counterfactual setting Y' , which gives premise (2). We now drop the simplification (of there being only two measurements) and return to Palmer’s full theory, allowing many measurements across the universe. Then his idea is that all such measurements determine a particular ontic state λ all at once. As before, this λ will not give outcomes for counterfactual settings, which again leads to premise (2).

While we agree that the ontic state λ is correlated with the setting choices X and Y , this appears to be because λ is essentially the measurement outcome, and it is natural to expect measurement outcomes to be correlated with measurement settings. To address SI, we need to look at the process by which a particular invariant set λ is determined using Alice’s and Bob’s free choices, and the ontic state λ_0 that existed ‘before’ λ . Palmer does not introduce λ_0 , but it is implicit in his model. This is because if Alice and Bob are genuinely free, then they could have measured otherwise, in which case, there would have to be facts about what outcomes they would get, had they chosen these alternative settings. Something (λ_0) must determine these counterfactual outcomes. As such, it is not clear that SI is violated by these measurements, with respect to the ontic state λ_0 , which exists prior to any measurements being chosen. Palmer is clear that Alice and Bob are completely free, meaning that the choice of measurement settings is not restricted by conditioning on λ_0 , i.e., $p(X, Y) = p(X, Y | \lambda_0)$.

What we think Palmer’s argument in (1)-(4) really shows is not a violation of SI, but a violation of a principle that is sometimes called *counterfactual realism*, where eigenvalues exist for all observables, even those that are not measured. This principle conflicts with quantum uncertainty principles, quantum contextuality Kochen and Specker (1967), and most directly the Leggett-Garg inequality (Leggett and Garg 1985), so its violation is not particularly surprising. On the upside, this means that Palmer’s 2024 theory, while nonlocal, is perfectly consistent with the scientific method, which may be a problem for genuinely superdeterministic theories (see section 3.3).

3 Philosophical consequences of superdeterminism

Having spelled out a broader space of possible superdeterministic theories, we now look to philosophical consequences, and try to resolve some disputes in the literature. In particular, we focus on the issues of free will (3.1), conspiracy (3.2), and scientific method (3.3). Each section is self-contained and can be read independently of the others.

3.1 Is superdeterminism compatible with free will?

Bell often described superdeterminism as a threat to free will. For example, in his final paper on his theorem (Bell 1990), he says,

“An essential element in the reasoning here is that [the measurement settings] are free variables. One can envisage then theories in which there just are no free variables [...]

In such ‘superdeterministic’ theories the apparent free will of experimenters, and any other apparent randomness, would be illusory.”

Bell has since been frequently accused of “causing confusion” with such remarks (Baas and Le Bihan 2023, Allori 2024), and it now seems widely accepted that superdeterminism has nothing to do with free will (see also Esfeld 2015). The following argument, from Baas and Le Bihan (2023), represents a commonly held view:

“superdeterminism is the conjunction of determinism and the atypicality of cosmological initial conditions [...] and as such, is no more problematic for free will than any deterministic theory. If one accepts the incompatibilist claim that free will is incompatible with determinism, it follows trivially that free will is incompatible with superdeterminism. In fact, there are many ways to reconcile free will with determinism, namely to endorse a form of compatibilism (for an overview, see e.g. McKenna and Coates [2021]). If free will is compatible with determinism, it is not clear why it should be incompatible with superdeterminism.”

The idea is that if superdeterminists simply embrace compatibilism, the view that free will is compatible with determinism, then they will be no worse off than ordinary determinists. Here, we argue that this is insufficient. For there are several prominent versions of compatibilism (described in McKenna and Coates 2021) that prevent experimenters in superdeterministic worlds from having the same sorts of freedoms as experimenters in ordinary deterministic worlds.

According to *classical compatibilism*, Alice’s choice to measure X was free if she could have chosen otherwise (e.g. by measuring Y). Some classical compatibilists tried to make this idea precise in conditional terms (Hume 1975, Ayer 1954, Hobart 1934). Thus, to say that Alice could have measured Y and not X is to say that, had she wanted (chosen, willed, or decided) to measure Y and not X at that time, then she would have measured Y and not X . In a merely deterministic theory, Alice’s alternative want is physically possible and would lead to that want being satisfied (given the different initial conditions, i.e., the different wants).

But in a superdeterministic theory, this may not be the case, especially in the context of nomic exclusion models. Consider the example described in Chen (2021), in which entangled photon pairs are either absorbed or transmitted by polarizers that are oriented in various different directions. Assume the left polarizer is set at 0 degrees, while the right polarizer is set at 30 degrees. Following Andreoletti and Vervoort (2022), call this “set-up A”, and call the actual photon ensemble measured by this set-up “sub-collection a ”. According to Andreoletti and Vervoort (2022), superdeterminism entails that if set-up B (where both polarizers are rotated another 30 degrees each) had instead been chosen, then a completely different photon ensemble (call it b) would have been measured. This is clear in goblin theory 3 ($X \rightarrow G \rightarrow \lambda$) from section 2.2.3, where the goblins select ensembles to be measured based on what measurement settings will be chosen. Now take the experimenter who chooses set-up A, and who therefore measures ensemble a . If the experimenter believes she could have measured *this* ensemble (sub-collection a) with set-up B, then she is just wrong. She is not free to do otherwise. So her choice of set-up A for sub-collection a was

not free. Similar reasoning applies to the other two goblin theories. Similar reasoning also applies to fine-tuned versions of superdeterminism, especially if the postulate about the initial conditions is a genuine law or constraint (Baas and Le Bihan 2023).

Note that the limitation on free will is restricted. The experimenter could not have chosen set-up B *for sub-collection a*, meaning she was not free according to classical compatibilism. But there are plenty of other things she is free to do. But this is a clear restriction on the kind of freedom we think we have, and it is a freedom that classical compatibilism will deliver for ordinary deterministic theories, where set-ups and sub-collections are independent. Superdeterminism - at least in the context of nomic exclusion models - therefore restricts our freedom more than ordinary determinism does, according to classical compatibilism.

Classical compatibilism has fallen out of favour among modern compatibilists, largely due to some apparent counterexamples to the conditional analysis (Chisholm 1964, van Inwagen 1983). If I have a crippling fear of snakes, then I cannot freely pick up a snake, even though, if I had wanted to pick one up, I would. Nonetheless, some modern versions of compatibilism still return a result that would vindicate Bell's concerns about free will. An example is what McKenna and Coates (2021) call *new dispositionalism*, advocated for example by Vihvelin (2004, 2013) and Smith et al. (2003).

According to the new dispositionalists, we hold fixed the relevant causal base or underlying structure of an agent's disposition to do something (like pick up a snake or implement set-up B on sub-collection *a*), and we consider various counterfactual conditions in which that causal base or underlying structure operates unimpaired. Does the agent in an appropriately rich range of such counterfactual conditions implement different set-ups on the same ensemble, or handle snakes? If so, then even if in the actual world she does not implement set-up B (but used A instead), or does not pick up a snake, she was able to do so. For she had at the time of action the pertinent agential abilities or capacities, even if the world is determined. But if a nomic exclusion version of superdeterminism is true, then the relevant counterfactual conditions do not allow one's disposition to implement set-up B on ensemble *a* to ever manifest. Thus, modern versions of compatibilism that are sensitive to the kinds of alternative possibilities that prominent superdeterministic models rule out, will support Bell's concerns about free will.

It should be noted that there are also many modern compatibilist views which disassociate free will from the ability to perform alternative actions completely. For example, Frankfurt (1971)'s *hierarchical mesh theory*, which explains freely willed action in terms of actions that issue from desires that mesh with hierarchically ordered elements in one's psychology. On this view, free actions stem from desires nested within more encompassing elements of oneself, such as second order desires, desires to have certain desires, or desires to be a certain sort of person. If my implementation of set-up A, came from desires within me, and about the type of rigorous experimenter I want to be, then I freely implemented that experiment, whether or not there are any goblins, or fine-tuned initial conditions. We therefore do not intend to claim that superdeterminism undermines free will. We only claim that it is an open question, which is not resolved merely by pointing to compatibilism.

To conclude, we consider a quite different argument for the independence of superdeterminism and free will from Allori (2024):

“More generally, one can see that the issue of free will is of no importance in these matters by observing that to define what SI is one needs no human beings. In fact, while it is the case that usually humans select the experimental settings, these could well be chosen by some sort of automatic random generator. If that is the case, then no genuine choice is involved or needed to define the hypothesis of SI.”

But this argument is not valid. We can similarly observe that to define what determinism is, one needs no human beings. But that does not stop determinism from being important to the issue of free will.

3.2 Is superdeterminism conspiratorial?

A common objection to superdeterminism is that it is *conspiratorial*. Perhaps the first to articulate this objection was Shimony, Horne, and Clauser (1976), who noted that

“In any scientific experiment in which two or more variables are supposed to be randomly selected, one can always conjecture that some factor in the overlap of the backwards light cones has controlled the presumably random choices. But, we maintain, skepticism of this sort will essentially dismiss all results of scientific experimentation. Unless we proceed under the assumption that hidden conspiracies of this sort do not occur, we have abandoned in advance the whole enterprise of discovering the laws of nature by experimentation.”

Similar points are pressed by Maudlin (2014), Sen and Valentini (2020b), Chen (2021), Allori (2024), and others. Meanwhile, authors sympathetic to superdeterminism have denied that superdeterminism is conspiratorial in any problematic sense.

Hossenfelder and Palmer (2020) interpret what they call “the conspiracy argument” as pointing out that the range of initial conditions that result in the relevant correlations between settings and λ are narrow. So to respond to it, they consider a theory with a constrained state space and argue that, within such a space, a physically relevant measure might reveal that the set of initial conditions yielding the correlations is broad. We do not find this response compelling. Such a theory will exclude possibilities which will *inevitably seem possible to observers*, such that according to the theory, nature is systematically deceiving us. This, as we will argue, is still conspiratorial.

More recently, Andreoletti and Vervoort (2022) have tried to rebut the conspiracy objection. They argue that the objection can be understood as the claim that superdeterminism entails the truth of “suspicious counterfactuals”. They consider the example described in Chen (2021), in which entangled photon pairs are either absorbed or transmitted by polarizers that are oriented in various different directions. Oriented in one way, quantum mechanics (correctly) predicts a certain fraction of photons will be absorbed. Oriented in another way, quantum mechanics predicts a different fraction. Assume the left polarizer is set at 0 degrees, while the right polarizer is set at 30

degrees. Andreoletti and Vervoort call this “set-up A”. They refer to the actual photon ensemble measured by this set-up as “sub-collection *a*”. They then note that according to superdeterminism, if set-up B (where both polarizers are rotated another 30 degrees each) had instead been chosen, then a completely different photon ensemble (call it *b*) would have been measured. Thus, superdeterminism apparently entails “suspicious counterfactuals” such as (C):

(C) If the set-up B had been chosen, then the sub-collection *b* would have been selected.

And this seems bizarre! For surely, if we had chosen set-up B instead of A, our ensemble would not have been different, we would still be experimenting on the same ensemble. Thus, Andreoletti and Vervoort argue that

“the charge of conspiracy against superdeterminism lies *precisely* in counterfactuals such as (C). That is, people who find superdeterminism a conspiracy theory and hence a non-starter do it because they think superdeterminism implies (C) and (C) is just hard to believe.”

Let us grant for the moment that this is what the conspiracy charge amounts to. Then, Andreoletti and Vervoort offer a response that is not unreasonable. They say that (C) is the wrong counterfactual to be focusing on. The correct one to consider is (C*): If the set-up B had been chosen *and nature is local*, then the sub-collection *b* would have been selected. So if we insist, as superdeterminists typically do, that nature is local, then the fact that sub-collection *b* would have been selected should not seem so mysterious.

However, as we have already suggested, this is not the right way to think about the conspiracy objection. To establish that a superdeterministic theory is conspiratorial, we need a clear and intuitive sufficient condition for a theory being “conspiratorial”. We then need to show that the superdeterministic theory satisfies the condition. We propose that a theory is conspiratorial if it entails that experimenters are *systematically deceived* by nature.

Whether it happens by fine-tuned initial conditions, statistical flukes, or nomic exclusion, superdeterminism implies that our experiences will lead us to believe that we are able to perform certain experiments on representative ensembles, when in fact we cannot.¹¹ For example, we are simply unable to use set-up B to measure sub-collection *a*. Yet, every experiment we’ve performed in the past has led us to believe that such a feat is entirely possible. For we have performed experiments just like *B* on ensembles that seem just like *a* in the past. So we are led to believe we can do it, but we never actually can. We think this is manifestly conspiratorial. Does this mean superdeterminism is unscientific? We will consider this in the next section.

¹¹Note that if these experiences systematically lead us to believe in libertarian free will, so that we “could have done otherwise” in a sense that’s incompatible with determinism, then any deterministic theory will seem conspiratorial by our definition. But belief in libertarian free will is hardly systematic. Given the prevalence of compatibilism, we cannot therefore say that given determinism, we are *deceived* into believing we have free will, because on the compatibilist view, we actually do have free will. Deterministic theories are therefore not in general conspiratorial, according to our definition.

3.3 Is superdeterminism unscientific?

The “objection from the impossibility of science” (as Andreoletti and Vervoort (2022) call it), has been posed by many authors, in a variety of ways, for example, Shimony, Horne, and Clauser (1976), Baas and Le Bihan (2023), Goldstein et al. (2011), Chen (2021), and Allori (2024).

Shimony, Horne, and Clauser (1976) argue that denying SI is “wrong on methodological grounds” if no specific causal linkage is proposed. For unless we “proceed under the assumption that hidden conspiracies of this sort do not occur, we have abandoned in advance the whole enterprise of discovering the laws of nature by experimentation”. Maudlin (2014) also argues that denying SI “would undercut scientific method.” The reason is that all “scientific interpretations of our observations presuppose that they have not have been manipulated in such a way.” For example, if every mouse that is exposed to cigarette smoke is found to get cancer, then we may inductively infer that cigarette smoke causes cancer in mice generally. But if our mice samples are not representative of mice (e.g. because goblins tamper with the lungs of our samples), then we cannot use induction at all, and science is ruined. Baas and Le Bihan (2023) go a step further, and argue that if SI were false (generally) in a theory, then it would suffer from the problem of *empirical incoherence*: its truth undermines our reasons to believe it (Barrett 1996). If the theory is true, then nature causes us to believe false theories.

Somehow, superdeterminists must justify the restriction of SI violation to Bell experiments. Chen (2021) concedes that “it is logically consistent for one to claim that statistical independence is false about microscopic systems but for all practical purposes true of macroscopic systems”. He considers decoherence as a possible mechanism for explaining how this could be, but concludes that it cannot offer any such explanation, and so suggests that there is a serious challenge here for explaining why SI violations aren’t ubiquitous.

Baas and Le Bihan (2023) introduce the term *exceptionalist SD* to refer to superdeterministic theories in which SI is obeyed in almost all circumstances, with the only exception being measurement settings for entangled quantum states, like those used in a Bell test. Most superdeterministic theories in the literature are intended to be exceptionalist, so that it remains safe to assume SI in all other areas of science. Because exceptionalist SD theories have cases where SI is obeyed, Baas and Le Bihan (2023) claim that “they must assert the impossibility of using those systems in order to set the measurement settings [in a Bell test].” Since it is otherwise unclear how one might ‘use those systems,’ we assume Baas and Le Bihan mean the systems used to perform the random selection procedures for cases that obey SI. An exceptionalist SD theory would then need to make the same random selection procedure impossible for choosing measurement settings in a Bell test. But we think this is incorrect. For example, we could flip coins and use the results of the same coins to choose settings in a Bell test and also to assign subjects to different groups in a drug trial, and this could result in representative samples for the subjects in the trial, while failing to produce representative samples for the Bell experiment. That is, in such a theory, we could have $\rho(\lambda_{\text{drug}}) = \rho(\lambda_{\text{drug}}|X)$ and $\rho(\lambda_{\text{Bell}}) \neq \rho(\lambda_{\text{Bell}}|X)$, where λ_{drug} are the ontic states that determine how a subject in a drug trial responds to the treatment they receive, and λ_{Bell} determine the out-

comes of measurements in a Bell experiment. So exceptionalist SD may be a logically consistent option, but given that the validity of the entire scientific method is at stake, it behooves proponents of exceptionalist SD, like Andreoletti and Vervoort (2022) and Nikolaev and Vervoort (2023), to provide the kind of “specific causal linkage” that Shimony, Horne, and Clauser (1976) saw was lacking, i.e., to explain and justify the exception to the SI rule.

What then are we to make of the scientific status of superdeterminism? Should we deem it unscientific or pseudo-scientific? Here we think it is instructive to consider the widely accepted framework of Lakatos (1970), which is designed to answer such questions. Lakatos urges us to think less in terms of particular theories, and more in terms of *research programmes*. A research programme is a sequence of falsifiable theories characterized by a shared core of assumptions that the programme takes for granted. For superdeterminism, these assumptions may include limited violations of SI and locality. A research programme then constitutes good science (science that is rational to work on and develop) if it is *progressive*, and bad science, if it is *degenerating*. Progressive programmes must meet two conditions. First, they must be theoretically progressive: each new theory in the sequence must have excess empirical content over its predecessor; it must predict novel and hitherto unexpected facts. Second, they must be empirically progressive. Some of that novel content has to be experimentally confirmed. If these conditions fail, the programme is degenerative.

The superdeterminism research programme cannot be described as progressive. For although some experimental tests for some superdeterministic models have been proposed (e.g. Hossenfelder 2011), they are not compelling. But it is also too early, we think, to consider superdeterminism inevitably degenerative. The mere logical possibility of a limited sort of statistical dependence in certain quantum experiments shows *there is no proof* that superdeterminism entails ubiquitous violations of SI. There is also no proof that we could not one day empirically discover the conspiratorial dependencies (e.g., we uncover the “goblins”).

Superdeterminism, presently, is therefore neither scientific nor unscientific, it is better thought of as being in a *pre-scientific* stage, where we are still coming to grips with what such models require, and we are only just beginning to propose simple toy models, several of which have been analyzed here. We therefore see no conclusive reason why the superdeterminism programme should be abandoned, as it may yet yield scientific fruit.

4 Conclusion

In this paper we have adopted and motivated the frequency interpretation of SI and we have argued that SI-violating superdeterministic models are not confined to deterministic models with atypical or fine-tuned initial conditions. We argued for two additional categories of superdeterminism, one involving fluke postulates, the other involving nomic exclusion, and we showed that several existing models fall into the latter category. We then argued that the attempt to categorize retrocausal models and invariant set theory as SI-violating is more subtle than has been appreciated, and depends upon exactly how we understand these models. Finally, we argued for three philosophical

consequences. First, some compatibilist accounts of free will entail that given superdeterminism, experimenters are not entirely free. Second, superdeterminism is conspiratorial in the sense that it postulates that nature is deceiving us. Third, superdeterminism is not unscientific but pre-scientific. We hope this motivates further research into the real meaning of SI-violation and the space of possible superdeterministic theories.

Acknowledgments: Thanks to Tim Palmer, Wayne Myrvold, and Michael Robinson, for several helpful discussions. This project/publication was made possible through the support of Grant 63209 from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

References

- Adlam, Emily C (2022). “Two roads to retrocausality”. In: *Synthese* 200.5, p. 422.
- Adlam, Emily C, Jonte R Hance, Sabine Hossenfelder, and Tim N Palmer (2024). “Taxonomy for physics beyond quantum mechanics”. In: *Proceedings A*. Vol. 480. 2294. The Royal Society, p. 20230779.
- Allori, Valeria (2024). “Hidden variables and Bell’s theorem: Local or not?” In: *THEORIA. An International Journal for Theory, History and Foundations of Science* 39.2, pp. 143–163.
- Andreoletti, Giacomo and Louis Vervoort (2022). “Superdeterminism: a reappraisal”. In: *Synthese* 200.5, p. 361.
- Ayer, A. J. (1954). “Freedom and Necessity”. In: *Philosophical Essays*. New York: St. Martin’s Press, pp. 3–20.
- Baas, Augustin and Baptiste Le Bihan (2023). “What does the world look like according to superdeterminism?” In: *The British Journal for the Philosophy of Science* 74.3, pp. 555–572.
- Barrett, Jeffrey A. (1996). “Empirical adequacy and the availability of reliable records in quantum mechanics”. In: *Philosophy of Science* 63.1, pp. 49–64.
- Bell, J.S. (1964). “On the Einstein Podolsky Rosen paradox”. In: *Physics* 1.3, p. 195.
- (1971). “Introduction to the Hidden-Variable Question”. In: *Foundations of Quantum Mechanics*. Ed. by B. d’Espagnat. New York: Academic Press, pp. 171–181.
- (1977). “Free variables and local causality”. In: *Epistemological Letters* 15. Reprinted in *Speakable and Unspeakable in Quantum Mechanics*, 1985, and in the 2nd edition, 2004, pp. 100–104, pp. 79–84.
- (1990). “La nouvelle cuisine”. In: *Between Science and Technology*. Ed. by A. Sarlemijn and P. Kroes. Elsevier Science Publishers.
- Bostrom, Nick (2003). “Are we living in a computer simulation?” In: *The philosophical quarterly* 53.211, pp. 243–255.
- Chalmers, David J. (2022). *Reality+: Virtual Worlds and the Problems of Philosophy*. New York: W. W. Norton.

- Chen, Eddy Keming (2021). “Bell’s Theorem, Quantum Probabilities, and Superdeterminism”. In: *The Routledge companion to philosophy of physics*. Routledge, pp. 184–199.
- Chisholm, Roderick (1964). *Human Freedom and the Self*. The Lindley Lectures. Lawrence, Kansas: Department of Philosophy, University of Kansas.
- Ciepielewski, Gerardo S, Elias Okon, and Daniel Sudarsky (2023). “On superdeterministic rejections of settings independence”. In: *The British Journal for the Philosophy of Science* 74.2, pp. 435–467.
- Clauser, John F. and Michael A. Horne (1974). “Experimental consequences of objective local theories”. In: *Physical Review D* 10, pp. 526–535.
- Costa de Beauregard, Olivier (1976). “Time symmetry and interpretation of quantum mechanics”. In: *Foundations of Physics* 6.5, pp. 539–559.
- Donadi, Sandro and Sabine Hossenfelder (2020). “A Superdeterministic Toy Model”. In: *arXiv preprint arXiv:2010.01327*.
- Dürr, Detlef and Stefan Teufel (2009). *Bohmian Mechanics: The Physics and Mathematics of Quantum Theory*. Berlin: Springer.
- Esfeld, Michael (2015). “Bell’s theorem and the issue of determinism and indeterminism”. In: *Foundations of Physics* 45.5, pp. 471–482.
- Frankfurt, Harry (1971). “Freedom of the Will and the Concept of a Person”. In: *Journal of Philosophy* 68, pp. 5–20.
- Goldstein, Sheldon, Travis Norsen, Daniel Victor Tausk, and Nino Zanghì (2011). “Bell’s theorem”. In: *Scholarpedia* 6.10, p. 8378. DOI: 10.4249/scholarpedia.8378.
- Hance, Jonte R (2024). “Counterfactual restrictions and Bell’s theorem”. In: *Journal of Physics Communications* 8.12, p. 122001.
- Hance, Jonte R and Sabine Hossenfelder (2022). “The wave function as a true ensemble”. In: *Proceedings of the Royal Society A* 478.2262, p. 20210705.
- Hance, Jonte R, Sabine Hossenfelder, and Tim N Palmer (2022). “Supermeasured: Violating Bell-Statistical Independence without violating physical statistical independence”. In: *Foundations of Physics* 52.4, p. 81.
- Hemmo, Meir and Itamar Pitowsky (2007). “Quantum Probability and Many Worlds”. In: *Studies in History and Philosophy of Modern Physics* 38.2, pp. 333–350.
- Hobart, R. E. (1934). “Free Will as Involving Indeterminism and Inconceivable Without It”. In: *Mind* 43, pp. 1–27.
- Hossenfelder, Sabine (2011). “Testing Super-Deterministic Hidden Variables Theories”. In: *Foundations of Physics* 41.9, pp. 1521–1531.
- (2020). “Superdeterminism: A guide for the perplexed”. In: *arXiv preprint arXiv:2010.01324*.
- Hossenfelder, Sabine and Tim N Palmer (2020). “Rethinking superdeterminism”. In: *Frontiers in Physics* 8, p. 139.
- Hume, David (1975). *An Enquiry Concerning Human Understanding*. Ed. by P.H. Nidditch. Oxford: Clarendon Press.

- Kochen, S. and EP Specker (1967). “The problem of hidden variables in quantum mechanics”. In: *J. of Math. and Mech.* 17, pp. 59–87.
- Lakatos, Imre (1970). “Falsification and the Methodology of Scientific Research Programmes’ in I. Lakatos and A. Musgrave (eds.) *Criticism and the Growth of Knowledge*”. In: *Proceedings of the International Colloquium in the Philosophy of Science*. Vol. 4. 91, p. 196.
- Leggett, Anthony J and Anupam Garg (1985). “Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks?” In: *Physical Review Letters* 54.9, p. 857.
- Leibniz, Gottfried Wilhelm (1714). *Monadology*. Ed. by G. H. R. Parkinson. Originally published in 1714. London: Everyman’s Library, pp. 179–194.
- Maudlin, Tim (2014). “What Bell did”. In: *Journal of Physics A: Mathematical and Theoretical* 47.42, p. 424010.
- McKenna, Michael and D. Justin Coates (2021). “Compatibilism”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2021. Metaphysics Research Lab, Stanford University.
- Myrvold, Wayne, Marco Genovese, and Abner Shimony (2020). “Bell’s Theorem”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2020. Metaphysics Research Lab, Stanford University.
- Nikolaev, Vitaly and Louis Vervoort (2023). “Aspects of superdeterminism made intuitive”. In: *Foundations of Physics* 53.1, p. 17.
- Palmer, Tim N (2022). “Discretised Hilbert Space and Superdeterminism”. In: *arXiv preprint arXiv:2204.05763*.
- (2024). “Superdeterminism without Conspiracy”. In: *Universe* 10.1, p. 47.
- Price, Huw and Ken Wharton (2015). “Disentangling the quantum world”. In: *Entropy* 17.11, pp. 7752–7767.
- Sen, Indrajit and Antony Valentini (2020a). “Superdeterministic hidden-variables models I: non-equilibrium and signalling”. In: *Proceedings of the Royal Society A* 476.2243, p. 20200212.
- (2020b). “Superdeterministic hidden-variables models II: conspiracy”. In: *Proceedings of the Royal Society A* 476.2243, p. 20200214.
- Shimony, Abner, Michael A. Horne, and John F. Clauser (1976). “Comment on ‘The theory of local beables’”. In: *Epistemological Letters* 13. Reprinted in Shimony, Horne, and Clauser (1985), and in Shimony (1993), 163–167, pp. 1–8.
- Smith, Michael et al. (2003). “Rational capacities, or: How to distinguish recklessness, weakness, and compulsion”. In: *Weakness of will and practical irrationality*, pp. 17–38.
- Sutherland, Roderick Ian (1983). “Bell’s theorem and backwards-in-time causality”. In: *International Journal of Theoretical Physics* 22, pp. 377–384.
- (2008). “Causally symmetric Bohm model”. In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 39.4, pp. 782–805.
- t’Hooft, Gerard (2016). *The cellular automaton interpretation of quantum mechanics*. Springer Nature.

- Vaidman, Lev (2016). “The Bell Inequality and The Many-Worlds Interpretation”. In: *Quantum Nonlocality and Reality: 50 Years of Bell’s Theorem*. Ed. by Mary Bell and Shan Gao. Cambridge University Press, pp. 195–203.
- van Inwagen, Peter (1983). *An Essay on Free Will*. Oxford: Clarendon Press.
- Vihvelin, Kadri (2004). “Free will demystified: A dispositional account”. In: *Philosophical Topics* 32.1/2, pp. 427–450.
- (2013). *Causes, laws, and free will: Why determinism doesn’t matter*. Oup Usa.
- Waegell, Mordecai and Kelvin J. McQueen (2020). “Reformulating Bell’s theorem: The search for a truly local quantum theory”. In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 70, pp. 39–50.
- Waegell, Mordecai, Kelvin J. McQueen, and Emily C Adlam (2023). “The Generative Programs Framework”. In: *arXiv preprint arXiv:2307.11282*.
- Wallace, David (2012). *The emergent multiverse: Quantum theory according to the Everett interpretation*. Oxford University Press.
- Wharton, Ken (2018). “A new class of retrocausal models”. In: *Entropy* 20.6, p. 410.
- Wharton, Ken and Nathan Argaman (2020). “Colloquium: Bell’s theorem and locally mediated reformulations of quantum mechanics”. In: *Reviews of Modern Physics* 92.2, p. 021002.
- Wiseman, Howard M and Eric G Cavalcanti (2017). “Causarum Investigatio and the two Bell’s theorems of John Bell”. In: *Quantum [Un] Speakables II: Half a Century of Bell’s Theorem*, pp. 119–142.