

Naturalness and Emergence

David Wallace

February 20, 2019

Abstract

I develop an account of naturalness (that is, approximately: lack of extreme fine-tuning) in physics which demonstrates that naturalness assumptions are not restricted to narrow cases in high-energy physics but are a ubiquitous part of inter-level relations are derived in physics. After exploring how and to what extent we might justify such assumptions on methodological grounds or through appeal to speculative future physics, I consider the apparent failure of naturalness in cosmology and in the Standard Model. I argue that any such naturalness failure threatens to undermine the entire structure of our understanding of inter-theoretic reduction, and so risks a much larger crisis in physics than is sometimes suggested; I briefly review some currently-popular strategies that might avoid that crisis.

1 Introduction

Physicists — more specifically, physicists working in ‘high-energy physics, i.e. particle physics and theoretical cosmology — have long made use of a criterion for their theories called ‘naturalness’, meaning roughly that the specification of a physical theory should not require extreme fine-tuning. But that criterion gives severely inaccurate results in several key situations — including the recent results from the Large Hadron Collider (LHC) at CERN. This has been met with consternation and with a renewed appraisal of the naturalness criterion: was it, after all, an unmotivated presupposition or even a mere aesthetic preference?

In fact, naturalness — properly understood — is a completely ubiquitous assumption right across the face of physics. It is unavoidable whenever we are trying to understand the relation between a higher-level, emergent physical theory and the lower-level physics from which it emerges — and although they concern energies and sizes incomprehensible to humans, the theories of particle physics and cosmology are themselves “higher-level, emergent”, arising from an as-yet unknown underlying physics.

This means that if naturalness really fails in high-energy physics, that failure undermines not just fine details of particle physics but the entire hard-won understanding we have of how physics describes systems at different levels and

how those descriptions interrelate with one another. The apparent failure of naturalness is then a crisis at the heart of contemporary physics.

In this article I will explain what I think is the correct understanding of naturalness and how it has applications far beyond high-energy physics. After a preliminary discussion (section 2) of the key distinctions between the form of a physical theory, the parameters (i.e. constants of nature) needed to specify it precisely, and the initial conditions fed into its equations, I begin (section 3) with the central role naturalness assumptions play in specifying the initial conditions of systems in statistical mechanics. In section 4 I extend the concept of naturalness from initial conditions to parameters; in section 5 I consider more generally the role naturalness conditions play in emergence, and how the success or failure of naturalness underpins a distinction close to, but more subtle than, the familiar distinction between strong and weak emergence in the literature.

Having established the significance of naturalness, in section 6 I turn to possible justifications of the naturalness assumption, and then (sections 7–8) discuss the apparent failure of naturalness in high-energy physics and its significance in the light of the previous discussion. Section 9 is the conclusion.

A note on terminology: ‘naturalness’ is used in the literature in widely different ways, ranging from narrow and technical to broad and sweeping. I make no attempt to adjudicate usage: my own use of the term is motivated by what seems the cleanest and most relevant way to describe the physics rather than from conformity to any particular precedent. For this reason I capitalise the term, referring throughout to Naturalness, to make clear that I intend to refer to the specific notion defined here. A note on sources: where the physics I discuss is new or speculative I give explicit citations, but I do not provide detailed or original sources for extant and well-known physics.

2 Specifying a physical system: equations, parameters, initial conditions

It is well known that the models constructed in physics decompose into two parts: the *dynamical equations* (the ‘laws’, if you like) that determine how one state evolves into another, and the *initial conditions* that specify what state the system actually starts in. But for our purposes it will be useful to operate with a rather finer-grained decomposition. For a system’s dynamical equations standardly depend on various coefficients, or to give them a grander name, on various constants of Nature. In electrodynamics, for instance, the repulsive force between two electrons famously is proportional to the product of their charges and inversely proportional to the distance between them — but only if we know the numerical value of the charge, and of the proportionality constant, are we in a position to actually calculate that force and to compare it with experiment.

Matters are complicated by the fact that the coefficients often have units, and these units are conventional. It is normal — at least in particle physics — to *define* that proportionality constant so that the inter-electron force is not

just proportional to, but *equal to*, e^2/r^2 , where e is the electron charge and r the inter-electron distance. That in turn guarantees that the charge of the electron has units of $\text{mass}^{1/2} \times \text{distance}^{3/2} \times \text{time}^{-1/2}$, but the actual value of the charge still depends on our choice of how to measure distance, time and mass.

The real physical content of a theory — insofar as that theory is considered on its own terms and not by reference to externally-understood standards of observation and measurement — is expressed by its *dimensionless coefficients*, which have no units and take the same value however we define our measurements. In (quantum) electrodynamics, for instance, the charge e of the electron, Planck’s constant \hbar , and the speed of light c all have units, but the *fine structure constant*,

$$\alpha \equiv e^2/4\pi\hbar c \simeq 1/137 \tag{1}$$

is dimensionless, does not depend on our choices of units, and can be seen as giving an objective measure of the strength of the electromagnetic interaction. In particle physics, for pure convenience we often just choose units in which $\hbar = c = 1$, so that they disappear from the theory and so that e can be replaced by $(4\pi\hbar c\alpha)^{1/2}$. Finding a dimensionless way to express the dimensionful parameters of a theory can be helpfully understood as finding some natural or useful way of defining the scales that characterise a problem or theory, and then expressing the parameters as multiples of those scales.)

In the following I will consider only dimensionless coefficients, setting aside some subtleties regarding dimensionful coefficients and assuming that they can be eliminated or reduced to dimensionless coefficients by some such appropriate choice of units.

We can now think of a model for some system as specified by three parts: the *qualitative* form of its dynamical equations (in which the coefficients are unspecified parameters); the actual, numerical values of the parameters; the initial condition. Only with all three can we make complete predictions for the system, although partial knowledge may suffice to make partial predictions. From one perspective, the parameters and the qualitative form of the equations go naturally together: both specify the lawlike features of the system, with the initial condition just picking out the contingent features. From another and more pragmatic perspective, it is only the qualitative form of the equations that we can even contemplate knowing exactly; the parameters and initial conditions can at most be known up to some margin of error.

In any case, my primary concern in this paper is the relation between different physical theories, and there is an obvious and intuitive way to think about it: given a higher-level theory that can be derived from a lower-level one, we might expect to derive the higher-level laws from the lower-level laws and the higher-level initial state from the lower-level initial state; and we might expect the qualitative form of the higher-level theory’s equations to be set by the qualitative form of the lower-level theory’s equations, with the lower-level parameters serving only to determine the higher-level parameters.

If we did expect this, we would be quite wrong: this ‘obvious and intuitive’ way of thinking is not in fact how inter-theoretic relations work in physics. As

we will see, even the qualitative structure of higher-level laws depends on some assumptions about the parameters and the initial conditions of the lower-level theory. To begin exploring this, let's consider the case of statistical mechanics, in which — we will see — the derivation of robust higher-level dynamics requires a subtle condition on the initial condition.

3 Initial-condition naturalness: the case of statistical mechanics

Suppose that we are interested in studying the evolution of a system of many components, under the assumption that it may be treated via classical mechanics,¹ and holding fixed for the moment the parameters that determine the exact form of the equations of classical mechanics. For instance, our system might be a box of gas, containing some 10^{23} particles. The system is deterministic: specifying the $3N$ positions and $3N$ momenta of all of the particles — that is, specifying a point in the $6N$ -dimensional phase space of the system (often called the system's *microstate*) — suffices to determine its evolution into the past and future. But for the most part, we are not concerned with the fine-grained details of the system's evolution but with some coarser-grained representation — the evolution of the system's spatial density, say, smeared over regions large compared to the interparticle distance but small compared to the size of the box. Discovering *coarse-grained* (or *collective*) *dynamics* for systems like this is the main business of non-equilibrium statistical mechanics.

There are many ways to carry out that business, but one of the conceptually clearest (if technically most cumbersome) is to partition the phase space into some large number of cells, each collecting together (say) all microstates of the system with the same coarse-grained particle mass and velocity density. (For more details of the framework I discuss here, see Wallace (2010, 2016), and references therein.) Each such cell is called a *macrostate*, and if in addition we introduce some interval τ of time, we can define a *coarse-grained history* of the system as a sequence of these macrostates, at successive times $t_0, t_0 + \tau, t_0 + 2\tau, \dots, t_0 + M\tau$. (Choosing the right coarse graining and time interval is as much art as science, and we will come back shortly to what makes any choice the 'right' one.) Any microstate of the system at initial time t_0 determines a macrostate for all subsequent times and so uniquely fixes a coarse-grained history; as a consequence we can partition the phase space into 'macrohistory regions', one corresponding to each coarse-grained history.

A coarse-grained *dynamics* — that is, a way of inferring the later macrostates in a history from the earlier ones, either deterministically or probabilistically — is somewhat harder to come by: in general, two fine-grained histories (that is, two dynamically-possible sequences of microstates) might share the same cell for a substantial period of time and then diverge, so that without some additional

¹The assumption can be lifted, with only slight conceptual consequences; but the details are technical, and for reasons of space I omit them.

input there is no prospect of an evolution rule at the macro-level. Micro-level features place *some* constraints on the coarse-grained histories, to be sure — conservation principles like the conservation of energy, in particular, restrict those coarse-grained histories that can be realised — but those constraints fall far short of specifying the actual coarse-grained dynamics which are frequently known empirically to hold for systems like this. And so if physicists are able to derive those dynamics from microscopic premises — and the practice of physics includes many such derivations — then there must be an additional ingredient.

That ingredient is usually understood probabilistically, since after all the coarse-grained dynamics we find in Nature are often probabilistic (think of Brownian motion, for instance). We can consider putting a specific probability measure over phase space and seeing what implications it has for the macrodynamics. A standard choice is the *Liouville measure*, which can be understood as in a certain sense uniform over phase space (more precisely, it is the unique measure that is invariant under any classically-permitted dynamics on phase space). That measure assigns a probability to each history, and hence a conditional probability to each history conditional on an initial subsequence of that history.

(What are these probabilities? Perhaps Bayesian quantifications of our ignorance of the true state (Jaynes 1957a, 1957b, Myrvold 2014; perhaps statements of relative frequency; perhaps some proxy for considerations of ‘typicality’ (Goldstein *et al* 2001); perhaps as Humean objective chances (Loewer 2002, Albert 2015, ch.1); perhaps as the classical limits of quantum-mechanical probabilities (Albert 2000, ch.7, Wallace 2016). For our purposes all that matters is that probability distributions of this kind are part of the core machinery of classical statistical mechanics.)

In a narrowly formal sense this defines a dynamics, in that it assigns a probability to any history, conditional on any initial segment of that history. But *a priori* it is entirely possible for that dynamics to be highly non-Markovian and highly time-dependent (that is, the probability of the system transitioning into a given macrostate at time $t + \tau$ can depend not only on the current macrostate but on all previous macrostates and on the exact value of t). But for the sorts of dynamical systems normally studied in statistical mechanics, in fact it turns out that the dynamics is much better behaved than that, defining (to a good approximation) a transition probability from one macrostate to another that depends only on the two macrostates. (Indeed, that a choice of macrostates and of time intervals leads to such well-defined transition probabilities is a large part of what makes a choice of macrostates a good one; cf Wallace (2012, ch.2) for further discussion.)

This might suggest that we should add ‘the initial probability measure is the Liouville measure’ as an axiom of statistical mechanics. But that axiom both delivers too much and requires too much. It defines not only *transition* probabilities but *absolute* probabilities: not just the probability that the system will later be found in macrostate Y given that it is currently in macrostate X , but the probability that it was found in macrostate X in the first place. And those absolute probabilities make the actually-observed state of many systems

stupendously unlikely: they give overwhelmingly dominant probability to a system being found in thermal equilibrium, whereas even cursory examination of our surroundings reveals many systems (ourselves not least) that are far from equilibrium. And it is also far stronger than is actually required to derive a dynamics: for exactly those systems where the Liouville probability delivers stable and well-behaved dynamics — and for exactly the same reason — those same dynamics are delivered by any probability distribution that is remotely smooth with respect to the Liouville measure.

The reason, in qualitative terms, is that the systems to which statistical-mechanical methods are applicable have strongly *mixing* properties, which means that the macrohistory regions we defined earlier are intermingled very closely together. For a given initial macrostate M , some portion p of its (Liouville) volume might correspond to a history h , and another portion p' to a history h' , so that the ratio of probabilities assigned to h and h' is p/p' . But the regions corresponding to the two histories are typically so irregular and so tangled with one another that if we pick pretty much any subregion of M and ask what the volume ratio of h and h' is when restricted to M , we still obtain p/p' . That means that our particular choice of the Liouville measure is largely irrelevant: almost any measure will deliver the same dynamics. (And in particular, the measure obtained by starting with the Liouville measure restricted to one macrostate and evolving it forward under the classical dynamics will deliver the same dynamics as the Liouville measure itself, which is why the macrodynamics are independent of the particular time index.)

‘Almost any’ measure, I said; but not *any* measure. After all, suppose that we pick a quite arbitrary set of transition probabilities for macrostates (except that we do not assign nonzero transition probabilities to transitions that are absolutely forbidden by the microdynamics, such as those which violate energy conservation), and a quite arbitrary initial probability distribution over macrostates, and use the two distributions to determine a probability distribution $P(h)$ over macrohistories h . Then if we choose a phase-space probability distribution that assigns probability $P(h)$ to set of initial microstates corresponding to history h (and there are infinitely many such distributions), *that* probability distribution will fail to deliver the standard Liouvillian transition probabilities but will instead return our arbitrarily-selected ones. Any such measure, however, will have an enormously complicated structure at the level of the phase-space geometry: it will vary very wildly with respect to the Liouville measure, being concentrated unevenly onto subsets of phase space whose own structure lacks a remotely simple description, other than by appeal to the macrodynamics themselves.

Anticipating the terminology of field theory, and writing μ for the Liouville measure, let us define a probability measure P on phase space as *Natural* if it has the form

$$P(X) = \int_X f(x) d\mu(x) \tag{2}$$

where f is some function on phase space specifiable in a reasonably simple way from the microscopic variables of the system. (The definition is unavoidably

imprecise; but I do need to require that our specification does not itself make use of dynamical notions such as ‘is the distribution obtained by evolving such-and-such distribution forward under such-and-such differential equation’.) Our results so far can then be summed up as: in those systems studied in non-equilibrium statistical mechanics, a stable classical macrodynamics is derivable from microdynamics under the additional assumption that the initial probability distribution is Natural, and that classical macrodynamics is independent of which Natural distribution is chosen. Let’s call that ‘stable classical macrodynamics’, in turn, the Natural macrodynamics.

(In earlier writing on the same subject (Wallace 2010), I called distributions like this *simple*. Somewhat similar ideas have been also advanced by Goldstein and co-workers under the label of *typicality*; cf Goldstein *et al* (2001), and also the discussion in Volchan 2007 and Frigg 2009, and references therein. The two concepts connect as follows: a property of a classical-mechanical system is Typical of a region R of phase space iff the subset S of states with that property is assigned conditional probability $P(S|R) \sim 1$ by any Natural probability measure on the system’s phase space, and Atypical iff that set is assigned conditional probability ~ 0 by any Natural measure.)

An ‘unNatural’ macrodynamics is any macrodynamics that is dynamically possible but which conflicts with the Natural macrodynamics. While initial distribution Naturalness is sufficient to derive the Natural macrodynamics, unNatural initial distributions do not generically lead to unNatural macrodynamics: given a randomly-selected unNatural measure, we would have no reason to expect that the ultra-complicated structure of its probability distribution would be just the *right* structure to give rise to some different macrodynamics. Quite the reverse, in fact: only very specific unNatural probability distributions encode stable anomalous macrodynamics, and most (in some admittedly vague sense of ‘most’) just succeed once again in picking out the normal macrodynamics. In particular, if we took some Natural distribution and applied some enormously complicated mathematical procedure to it, that procedure would scramble it in some way that might defy simple description, but we would have no reason to expect that it would scramble it in just the right way to disrupt the Natural macrodynamics and pick out some unNatural macrodynamics.

It is then helpful to distinguish between two classes of unNatural distributions. ‘Weakly unNatural’ distributions have a complicated structure that is not straightforward to describe, but nonetheless entail the Natural macrodynamics. ‘Strongly unNatural’ distributions entail some unNatural macrodynamics.

As a concrete application of these distinctions, suppose that we begin with a statistical-mechanical system (say, our dilute gas again) and a Natural probability distribution $\rho(0)$ over that system, suppose that the system’s initial macrostate is far from thermal equilibrium, and allow it to evolve forward for some time t . *Ex hypothesi* the system will obey the Natural macrodynamics over that period, but the probability distribution will be scattered very finely over the state space, so that the new probability distribution $\rho(t)$ is far from Natural. We still expect further evolution from time t to time $2t$ to conform to the Natural macrodynamics, though, so that $\rho(t)$ is only weakly unNatural.

However, if we apply the time-reversal operation τ to the distribution, reversing the velocities of all the particles, then the resultant distribution $\tau\rho(t)$ will evolve over a further time t to $\tau\rho(0)$. This will be in violation of the macrodynamics, demonstrating that $\tau\rho(t)$ is strongly unNatural: it is concentrated on exactly the right regions of phase space for the system’s past evolution to carry it back to its initial macrocondition.

4 Naturalness in parameter space: the case of effective field theory

So far we have been supposing the parameters of our system’s dynamics to be fixed, but there is a straightforward way to move the assumption of Naturalness from the initial condition to the parameters. Recall that only certain very precisely structured, highly unNatural probability distributions block the derivation of normal macrodynamics; that ‘precise structure’ depends on the value of the parameters, and will normally be disrupted by even very small changes of those parameters. So: suppose we represent the finite precision with which we might know the physical constants by some probability distribution over parameter space. If that distribution is Natural in the same sense as the initial-condition distribution — that is, it is specified by some not-ridiculously-complicated function, relative to the uniform distribution — then for a quite arbitrary (but fixed, parameter-independent) choice of initial-condition distribution, we have very good (albeit, as usual, heuristic) grounds to think that with extremely high probability the normal macrodynamics will be derivable. Only a very specific combination of initial conditions and parameters will give rise to anomalous macrodynamics.

The role of naturalness assumptions for parameters goes well beyond this case, however. In the rest of the paper, I will primarily be concerned with that role for one particular class of physical theories: *quantum field theories* (QFTs), which can be thought of as the most general form of quantum theory applicable to describe spatially extended bodies.² The best known examples of these are the eponymous *quantum fields* that seem to describe matter on the shortest lengthscales we know: the Standard Model of particle physics, for instance, is a QFT. But they are equally applicable to more mundane systems like crystals or iron bars (the subjects of ‘condensed-matter physics’), and it will be helpful to focus on these latter applications of QFT for the moment. (Apart from anything else, they make the point that QFTs are in no way confined to so-called ‘fundamental’ physics.)

Given some system that we want to model with a QFT, there is a fairly standard procedure, often called the ‘effective field theory’ method (the *term* is most commonly encountered in particle physics; the *idea* is common to particle

²For a more detailed discussion of the physics here, and references into the primary physics literature, see Wallace (2017); for an illuminating, but technical, discussion, see Polchinski (1999).

and condensed-matter physics). It goes like this: firstly, we identify a *cutoff scale*, a lengthscale below which the theory is not expected to be applicable. (In condensed matter physics, the ultimate cutoff comes from the fact that matter is not really continuous but is made of atoms.) The cutoff sets a fundamental scale for the theory; coefficients are usually rendered dimensionless by expressing them as multiples of some power of the cutoff.

Secondly, we write down the most general possible qualitative form of the dynamical equations compatible with the symmetries of the theory (which are postulated, determined experimentally, or some combination). Normally, that ‘most general possible qualitative form’ will require *infinitely* many dimensionless coefficients to specify a particular choice of dynamics, and that might suggest that the theory is devoid of predictive power: after all, we will ultimately have to determine the value of those coefficients through experiment, and if there are infinitely many coefficients to determine, it would seem that we would need infinitely many experiments before the theory ceases to be adjustable to fit the data. Surprisingly, this turns out not to be the case.

Thirdly, we ask what predictions we can make from the theory *if we confine ourselves to large-scale phenomena*, by which we mean: phenomena on lengthscales much longer than the cutoff scale. (In condensed matter physics, the cutoff scale is about 10^{-10}m — one ten-billionth of a meter — and so even micrometer-scale phenomena count as ‘large-scale’ by these lights.) The conceptually enlightening way to ask this question is to suppose that we understand the physics of phenomena at some lengthscale L , and ask how they are related to phenomena at a different scale kL . It turns out that (up to an overall rescaling) the physics of a given QFT at scale kL is identical to that of a different QFT at the original scale L — different in the sense that the values of the parameters have changed. In doing so, we can relate the study of a single QFT at a sequence of varying lengthscales to the study of a family of QFTs with varying parameters, all at a fixed lengthscale. This effective ‘flow’ through parameter space as we study the theory at increasing lengths is known as *renormalisation group flow*, and is one of the key insights of late-20th-century theoretical physics. It is not entirely obvious from the description I have given, but this is effectively another form of coarse-graining, just as in the move from micro- to macro-state in statistical mechanics: renormalisation group flow works by coarse-graining over the short-distance degrees of freedom and then rescaling the system.

For our purposes, the crucial thing about renormalisation group flow is that the parameters can be divided into three categories (the names of the categories are somewhat misleading and should not be taken too seriously):

- *Irrelevant parameters* rapidly decrease in size as the system flows to larger scales. Normally, all but a finite number of the parameters are irrelevant.
- *Relevant parameters* rapidly increase in size as the system flows to larger scales.
- *Marginal parameters* remain approximately constant in size at different points on the renormalisation group flow.

On this basis, physicists using QFT normally assume

- (i) that if the theory has any relevant parameters, those parameters will so dominate the dynamics as to violate the approximations and assumptions that justified using the QFT in the first place, so that QFTs with relevant parameters are not generally suitable for modelling physical systems. (Or put another way: if you made a hypothesis about a system that leads to it being modelled by a QFT with relevant parameters, rethink that hypothesis.)
- (i) that the irrelevant parameters will be so small as to be negligible (so that the dynamics of the theory is controlled by the marginal parameters).

But neither assumption *logically* follows from the renormalisation-group analysis. Substantive assumptions about the original ('bare') values of the parameters are required, and it will be important for us to consider what they are.

Schematically, the dependency of a parameter l_{cut} on the scale L looks like

$$\lambda(L) \sim \lambda_0 \times (l_{cut}/L)^n + C(L) \times (l_{cut}/L)^n \quad (3)$$

where l_{cut} is the cutoff length, λ_0 is the original value of the parameter, and $C(L)$ is another dimensionless parameter (calculated in a complicated way from the dynamics, and generally of order unity³) that does not depend too sensitively on L . We have $n > 0$ for irrelevant parameters, $n < 0$ for relevant parameters, and $n = 0$ for marginal parameters.

We can see that the effective value of an irrelevant parameter will be very small for $L \gg l_{cut}$ unless its original value is of order L/l_{cut} , which will be a very large dimensionless number. Physicists typically make what we might call the 'order one hypothesis' (or $O(1)$ hypothesis, in mathematical language), which is the hypothesis that dimensionless parameters that appear in theories are within a few orders of magnitude of unity. The rationale for this hypothesis is rarely spelled out explicitly (the clearest discussion I am aware of is in Barrow and Tipler (1986, pp.258-287)) but it seems some combination of the fact that dimensionless quantities appearing in fundamental physics rarely seem too large or too small, with the observation that the mathematical processes used in physics rarely seem to generate really large or really small factors. The $O(1)$ hypothesis is sometimes called 'naturalness' in the physics literature, but I will avoid that usage here; as we will see, it is a quite different phenomenon from true Naturalness.⁴

One might imagine that the converse would hold for relevant parameters: that they are very large unless their original value is very small. This is not the case: we can see from equation (3) that even if λ_0 is exactly zero, $\lambda(L)$ will still be very large. The only way for $\lambda(L)$ *not* to be very large is for λ_0 and $C(L)$

³I simplify here, though not in a way that essentially alters the main point: C depends on the values of all of the other parameters.

⁴Williams (2018) is an insightful recent discussion of the difference between these two conceptions of naturalness, though he draws the distinction somewhat differently than I do here.

to cancel out extremely precisely, to one part in $\sim (L/l_{cut})^n$. Even a very tiny perturbation in the value of λ_0 will cause the effective value of the parameter to become large.

The renormalisation-group strategy for analysing QFTs, then, makes two assumptions: that dimensionless parameters are not too large or small, and that they do not have the very precisely-tuned values they would need to have in order for any relevant parameters not to dominate the physics — and it is this second assumption that is most commonly called naturalness in the QFT literature. To understand this further, suppose that we again introduce a probability measure over the space of parameters — at least pro tem we can think of this measure as a Bayesian prior, with the caveat that we are using it as a helpful model of how to reason about QFTs, not as any uniquely-given rationality principle. We can again call any such measure Natural if it is specifiable in a tractable way starting with the uniform distribution over parameter space.

This ‘uniform measure’ is in fact significantly less well-behaved than the uniform Liouville measure of classical mechanics. For one thing, we lack a dynamical rule to specify it uniquely (recall that the Liouville measure is the unique measure invariant under dynamical flow for an arbitrary choice of dynamics). This can be seen vividly when we consider that we could always have parameterised a theory in terms of the squared value of the constants, or in terms of their exponents or logarithms. Secondly, since the parameters are not bounded in value, most plausible choices of uniform measure are not *probability* measures: they assign measure $+\infty$ to the entire space.

It is important to understand, though, that these subtleties cause no problems for the definition of Naturalness. A Natural measure over parameter space is any probability measure that is tractably specifiable in terms of the uniform measure, and each of the choices of uniform measure we have mentioned are Natural with respect to one another. So the consequence we wish to derive from the Naturalness assumption — that relevant parameters, if there are any, will blow up so as to dominate the long-distance physics — is highly insensitive to how we wish to precisify that assumption. We can say, consistently, that for any Natural measure it is extremely unlikely that the relevant parameters will not take very large values. Note by contrast that we *cannot* say that the $O(1)$ hypothesis is likely to hold for any Natural measure: that assumption is in an important sense much stronger than Naturalness, something which has not always been appreciated in the foundational literature.

Adapting the terminology of the previous section, let’s say that the qualitative form of the large-scale physics derived from a QFT is Natural if it arises from a Natural distribution over the short-distance parameter space. Two different Natural distributions will lead to large-scale dynamics with different values of parameters but the same qualitative form. The large-scale physics is unNatural if it does not so arise. Generic unNatural distributions also give rise to the Natural large-scale physics, and so we can call them weakly unNatural by analogy with the initial-condition case; only very carefully defined, ‘strongly unNatural’, distributions give rise to unNatural large-scale physics.

5 Emergence and its relation with Naturalness

The relation between the micro- and macrodynamics of a system falls under the vexed topic of *emergence*. In the literature on that topic, one distinguishes between *strong emergence*, in which the macrodynamics cannot even in principle be derived from microscopic considerations, and *weak emergence*, where such a derivation is at least in principle possible. In the case of strong emergence there is further ambiguity as to whether the macroscopic laws are underivable from the microscopic *laws*, or (more strikingly) underivable even from the microscopic laws and the microscopic states together. An abiding theme of this literature is how to advance a notion of emergence strong enough to encompass the widely-held view that macroscopic phenomena are in some sense novel and not mere redescriptions of the microscopic, without setting up a flat contradiction between micro and macro. (See Chalmers (2008), and references therein, for a fuller introduction to these concepts.)

Our analysis so far provides useful nuance to these distinctions, and clarifies the different roles played by assumptions about the qualitative form of the laws, the parameters, and the initial state. From the form of the microlaws alone there is no prospect of deriving any macrodynamics. But we need add to those microlaws only the (arguably!) minimal assumption of Naturalness in order to derive the macrolaws. Yet Naturalness is by no means logically entailed by the microdynamics, and we have seen that at the price of an unNatural initial condition and/or choice of probability on parameter space, we can construct virtually any macrolaws we like without violating the microdynamics.

So at least in the context of physics, it is natural to distinguish not exactly between strong and weak emergence, but between Natural and unNatural emergence — where a macrodynamics is Naturally emergent if it can be derived from the microdynamics together with any Natural probability distribution over initial conditions, and unNaturally emergent if it is compatible with the microdynamics but only by making an unNatural choice of initial distribution. I suggest that this distinction captures the spirit of the strong/weak distinction (again, at least as far as physics is concerned) better than the original.⁵

It is helpful to imagine two scenarios for how the world could be. In both scenarios, we can analyse the physics of many different systems on many different scales, and find autonomous dynamical equations for those systems that are not simple, mechanical applications of the same set of laws. But in both scenarios, descriptions of the same system on different scales, and of a system and its subsystems, are compatible with one another. We can define coarse-graining maps from the state space on one scale to the state space on another, so that the action of the coarse-graining map commutes with evolution under the respective systems' dynamics for the actual initial conditions of the systems in question.

⁵Ultimately, technical terms mean what we define them to mean, and I don't want to say that this is the *right* way to define strong and weak emergence, but rather that this dichotomy seems to do the work that the strong/weak distinction seems intended to do, at least in physics contexts. (It is probably much less well suited to distinguish the sense in which many philosophers argue that phenomenal consciousness is strongly emergent.)

In the first world — the ‘Naturally emergent’ world — something much stronger can be said about inter-theoretic relations. It is not just that a higher-level system’s evolution is compatible with a description of that system at a finer grain: the higher-level system’s dynamics can be *derived* from those of the lower-level system by making an additional assumption of Naturalness. Given a box of gas, say, we can understand its expansions and contractions by deriving the equations that govern them from classical particle mechanics under the Naturalness assumption. In such a world, and assuming that Naturalness can somehow be justified as an assumption, a form of reductionism holds: not a crude reductionism where the language and concepts of the higher level are simply dispensible in favor of a unified low-level description, but a sophisticated (and much more scientifically plausible) reductionism where the dynamical description of a given system on one level are grounded in and explainable in terms of the dynamical description of that same system on a lower level. Essentially all of the information about the world’s dynamics is encoded in the lowest-level dynamics, extractable from those dynamics through the assumption of Naturalness.

In the second world — the ‘unNaturally emergent’ world — things are very different. Attempting to derive higher-level physics in this way gives the wrong answer: the transition probabilities of the high-level dynamics systematically deviate from what would Naturally be expected, even though no transition occurs that is flatly forbidden by the lower-level physics, and the parameters of the higher-level equations are frequently different from what a renormalisation-group analysis would Naturally imply. In this world, systems’ initial conditions, and the appropriate choices of parameters, are strongly unNatural, in just the right way to generate the correct dynamics at each level. In this world, reductionism is false in any meaningful sense even though everything supervenes on the lowest-level facts. The vast majority of the information about the world’s dynamics is encoded in intricate details of the initial condition and of the precise values of the parameters.

Which world is closer to ours? It is an empirical question, and a contested one. The great majority of scientists (and the vast majority of physicists) adhere to something like the Naturally emergent picture of the world, even as many (rightly) scorn over-simplistic forms of reduction (Weinberg (1987) and Anderson (1972), for instance, famously disagreed about the extent to which particle physics provides a foundation for all of physics, but I think both would accept the Naturally emergent viewpoint). For powerful recent defenses of an appropriately sophisticated reductionism see Carroll (2017) and Dennett (1995, ch.3). But there are philosophers (e.g. Cartwright (1983, 1999)) and scientists (e.g. Ellis (2012a, 2012b)) who systematically reject the idea and regard the claim that the higher level is derivable from the lower (with or without naturalness) as false, or at least unestablished, in many or most scientifically interesting cases. (Cartwright, for instance, regards so-called ‘fundamental’ physics as at most a framework in which models at different levels can be constructed, and sees the evidence for real intertheoretic reduction as overstated at best; her metaphysics is very close to the unNaturally-emergent world).

The debate is far too large to settle here, but I will assume in this paper that the majority view in science is correct: at least in physics, and at least in most cases, we have pretty strong evidence that higher-level and lower-level dynamics are related in something like the naturally-Emergent fashion. My case for doing so, in brief, is: (a) physicists have constructed hundreds of inter-theoretic derivations of this kind, constructing higher-level dynamics from lower-level dynamics through naturalness assumptions, from fluid dynamics to plasma dispersion equations to superconductivity to Brownian motion; (b) while to be sure those constructions often have mathematically and conceptually shaky steps, we can gain some confidence in their underlying reliability by the fact that they so frequently predict not just the qualitative form of the higher-level equations but the numerical values of their parameters (as functions of the lower-level parameter values), something that would be an inexplicable miracle if there was not a mathematically valid derivation in the vicinity.

Repeated success in the construction of inter-theoretic relations does not, of course, mean that it is *always* possible: any example of a higher-level theory *not* derivable (with or without Naturalness assumptions) from the lower level would tell us that the Naturally emergent description is not unrestrictedly correct. Two specific examples are repeatedly raised in the emergence literature (see, e.g., Chalmers (2008) and Ellis (2012a)): phenomenal consciousness, and the quantum measurement problem.

Regarding consciousness, various thought experiments (see Chalmers (1996) and references therein) have convinced a great many philosophers that subjective experience is in principle irreducible to any dynamical process;⁶ it is often held up as the strongest plausible example of strong emergence, being undervivable even from the full listing of facts about the microworld (that is, not just the laws, but the full description of the physical state). But whatever the virtues of this case, it is *sui generis*; almost all advocates of irreducible consciousness accept that it is epiphenomal, having in turn no dynamical effects on the physical world. So I set it aside here.

The quantum measurement problem seems on the face of it to offer a better case of genuinely *dynamical* irreducibility: as any undergraduate physicist learns, ‘measurement causes the quantum wave function to collapse’, and ‘measurement’ is a higher-level notion admitting of no straightforward description in microphysical terms. But in the last few decades it has become extremely widely accepted that as long as we want to calculate empirical facts about how a measurement proceeds, the dynamical process of decoherence — par excellence a Naturally emergent process derivable from the underlying quantum dynamics — is entirely sufficient, with no need to pause the system’s normal dynamics in order to insert a collapse process. It remains contested whether decoherence entirely resolves the *conceptual* problems associated with quantum measurement (my own view, developed in detail in Wallace (2012) is that it does, but that taking it seriously pushes us towards the Everett - Many-Worlds - inter-

⁶While leaving many other philosophers, including the present author, cold; see Dennett (2001) for one of the better counter-arguments I know.

pretation of quantum mechanics). But at any rate there seems no evidence of any *dynamical* unNaturalness in the measurement process, and those physicists and philosophers who view the conceptual problems as reason to change the dynamics (such as advocates of the de Broglie-Bohm⁷ and GRW⁸ theories) almost all advocate changing it in a way that preserves a Natural story about intertheoretic relations.

So: let's assume that contemporary physics — at least putting aside the troubling cases coming from recent particle physics and cosmology, to which I return in section 7) is entirely consistent with the Naturally emergent picture. It then seems urgent to ask just why that picture is entitled to assume Naturalness, without which none of its supposed 'derivations' of high-level from low-level dynamics actually go through.

6 Justifying Naturalness

Is Naturalness just a reasonable methodological assumption, not logically derivable from the microphysics but justified on general principles of reasoning? The name conjures up the idea, and the apparent silliness of positing an extremely finely structured initial probability distribution supports it. Advocates of the closely-related notion of typicality (e.g., Goldstein *et al* 2001, Maudlin 2007) seem to discuss it in these terms, as a general aspect of scientific rationality; defenders of a Bayesian conception of probability like Myrvold (2014) say similar things. Myrvold in particular argues that a physically-realistic Bayesian agent could not make use of a probability distribution on phase space that was not at least somewhat simple. The naturalness(!) of Naturalness is displayed by the fact that its assumption looks so reasonable in textbook discussions of statistical mechanics that students⁹ often miss the fact that a substantive assumption has been made at all.

But that very example makes me cautious of an epistemological justification of Naturalness. As I noted briefly in section 3, dynamical evolution tends to turn Natural distributions into unNatural ones, with just the right form that their time reverse will display unNatural macrodynamics. And so insofar as our principles of inference do not themselves build in some assumed direction of time, it cannot simply be a principle of inference that a system's boundary conditions should be assumed Natural. (It is perhaps unsurprising to find that both Maudlin and Myrvold have philosophies of time and causation that are explicitly time-asymmetric, in contrast with the widespread view among physicists that any such asymmetry must be derivative on the direction of time in emergent macrophysics and not presupposed by the derivation of any such macrophysics.)

⁷Bohm (1952); for a contemporary presentation see Dürr, Goldstein, and Zanghi (1996).

⁸Ghirardi, Rimini, and Weber (1986); for a review of this and related theories, see Bassi and Ghirardi (2003).

⁹Textbook writers too, in some cases, even if it would be invidious to name names.

Arguments for Naturalness based on some epistemic reason to prefer simply-specified probability distributions also presume that we should give priority to microphysical descriptions of a system, rather than macro-level ones. After all, given a macrodynamics it is not difficult to specify that class of probability distributions that give rise to the macrodynamics: I just did it! The difficulty comes only when we try to specify it in terms of the microphysical degrees of freedom, without appeal to locutions like ‘gives rise to such-and-such macrodynamics’. But why assume that descriptions in terms of the macrophysics are unreasonable? We cannot appeal to the fact that macrophysics is underpinned by and derivable from microphysics, because the possibility of such derivations is exactly what is in question here. Methodological arguments for Naturalness seem to beg the question in favor of exactly the reductionist picture of science that is in question.

Instead of looking to the scientific method to justify Naturalness, we might look to physics itself. After all, the initial conditions and parameters of almost all systems in the Universe should in principle be derivable from the physics of other systems, where those ‘other systems’ are temporally earlier and/or describing more microscopic degrees of freedom. (For instance, the initial conditions of a classical theory of the dilute gas might be derivable from the initial conditions of a quantum theory of that same gas, or from a broader dynamical theory that models the way in which the box was prepared in the first place.) In this situation we have *three* dynamical systems to consider:

1. The *intermediate system*, which is the system whose initial conditions and/or parameters we are considering;
2. The *macro system*, whose macrodynamics we can derive from the intermediate system if we additionally make a Naturalness assumption;
3. The *precursor system*, from which we might hope to derive the Naturalness assumption for the intermediate system’s initial conditions.

In this situation, a Naturalness assumption for the precursor system can normally be expected to ground Natural emergence not just of the intermediate from the precursor, but of the macro from the intermediate. We should not expect to derive Naturalness of initial conditions or parameters for the intermediate system — the derivation may be complicated, and so the intermediate system’s parameters and initial conditions will be related to the precursor system by complicated mathematical expressions that will probably defy exact description. But — barring some remarkable coincidence in the mathematical forms of the theory — we would expect this to give rise to weakly Unnatural conditions for the intermediate system, and thus to continue underpinning Natural macrodynamics.

As such, once we have made the Naturalness assumption for a system at any given scale and any given time, there is no independent need to make it later and at larger scales, and so we can push the justification of Naturalness from higher-level to lower-level theories, as well as earlier and earlier in time. Does

this process end? This depends on features of our physics that at this stage are speculative, but at least the following factors may come into play as we go earlier and deeper:

Quantum origins of initial-condition probabilities: The classical theories we considered in section 3 were deterministic, and so any probabilities that occurred in the macrodynamics had to arise from probability assumptions over initial conditions. But the path to more fundamental physics inevitably leads us to quantum theory, in which probability is normally taken to be part of the dynamics and not simply imposed on initial conditions. This in turn suggests that the probability distributions we place over initial conditions (whose interpretation I left vague in section 3 are ultimately quantum-mechanical in origin, and more detailed investigation of the mathematics supports this hypothesis (Wallace 2016).

At this point it matters how we understand quantum theory itself. If some kind of deterministic hidden-variable theory (like the de Broglie-Bohm theory) is true, then there is no real change in the story: classical-statistical-mechanical probabilities are replaced by quantum probabilities, but these in turn are underpinned by a sort of statistical-mechanical probability applied to the hidden variables. If some theory of dynamical collapse (like the GRW theory) is true, then the dynamics can be expected to impose Naturalness dynamically, irrespective of a quantum system's initial state, because the stochastic dynamics can be expected to randomise the state and destroy the fine-grained details characteristic of strongly unNatural distributions. (Albert (2000) proposes, in effect, that the Naturalness condition can be understood this way, though he puts it rather differently.)

If — as is more or less the orthodoxy in modern high-energy physics and cosmology — we do not modify or supplement quantum theory in either of these ways, matters are a bit subtler: we can derive classical stochastic dynamics by means of decoherence theory, but that theory relies on the initial quantum state being Natural, in the sense that its structure is reasonably simply specifiable in terms of the microphysical structure of the theory. (For instance, in the quantum field theories used to describe the early Universe, we normally assume the initial state is the so-called ‘Bunch-Davies vacuum’, which has a very simple description; in more mundane examples in quantum statistical mechanics, we normally assume an initial state where system and environment are unentangled and separately admit of simple description. See Wallace (2010) for more.) In this case, we still need a Naturalness assumption on initial states; the difference is that the assumption no longer need be explicitly probabilistic.

Contingency of physical parameters: It might happen that what appears as a parameter (a lawlike feature) of a higher-level system is in fact a contingent matter from the point of view of the lower-level system. A mundane example: in constructing a model of the Solar system in which the planets are treated as point particles we normally take the masses of

the planets as parameters, invariant from model to model; yet of course those masses are contingent matters from the point of view of a deeper and more detailed model. If this occurs, it moves our Naturalness assumption from parameters to initial conditions (and in doing so allows for non-epistemic interpretations of the probability distribution over parameters in the first place).

Rigidity: As I noted in section 2, the theories we study in physics have a space of parameters and a space of initial conditions, and both can be chosen freely and varied smoothly. (Mathematically speaking, parameter space and state space both have the topology of some real manifold.) But speculatively, some future physical theory might lack this freedom. Physicists have long dreamed of being able to calculate the parameters in a theory from first principles; more recently, cosmologists have speculated that some unique principle might fix a system’s initial state, or even that the distinction between initial state and dynamics might dissolve in a quantum theory of gravity. In this situation, then — again, barring some remarkable coincidence in the mathematical forms of the theory — we would expect the emergent initial state and parameters of the higher-level theory to be Natural, or at worst weakly unNatural. We can distinguish two possibilities: *parameter rigidity* (the theory has no freely variable parameters) and *state rigidity* (the theory’s initial state, or equivalent, is fixed by some unique condition).

These ideas are illustrated by quantum gravity (albeit, those illustrations concern speculative physics that is at present far from empirical confirmation). String theory — by far the most thoroughly studied candidate for a quantum theory of gravity — is a quantum theory and so provides a quantum origin for probabilities; it has no dimensionless free parameters and so is parameter-rigid; the emergence of low-energy field theories from string theory has the parameters of those low-energy theories determined quantum-mechanically randomly and so the theory induces parameter contingency at the appropriate level.¹⁰ So if string theory is the true fundamental theory of the world, it plausibly provides an explanation for parameter Naturalness. However, string theory — at least as we currently understand it — is not state-rigid, so a string-theoretic cosmology would still require a Naturalness assumption for the initial state. (In another (compatible) corner of quantum gravity, the ‘no boundary’ proposal of Hartle and Hawking (1983) provides (in effect) a unique condition for the initial state of the Universe: on Hartle and Hawking’s proposal, quantum gravity would be state rigid.)

Where does this all leave us? Naturalness assumptions are consistent between theories at different levels, not in the sense that Naturalness at one level implies Naturalness at a higher level but in the still-sufficient sense that it implies weak unNaturalness, and so leaves Natural emergence intact. So insofar as

¹⁰See Tong (2009), and references therein, for an introduction to these ideas; for a non-technical discussion, try Greene (1999).

the Naturally emergent view is correct, we can impose Naturalness at as short a lengthscale and at as early a time as we can, and that imposition will suffice to justify the use of Naturalness assumptions at later times and higher levels of description. We are not yet in a position to know whether the assumption is a bare posit of fundamental physics, or follows from an underlying Rigid fundamental physics. But even if it needs to be a bare posit: that posit seems empirically justified by the (*ex hypothesi*) success of the Natural emergence account itself.

All of which seems to provide the substructure for a thoroughly intellectually satisfying, appropriately sophisticated, story about reduction and emergence. And all of which may be fatally undermined by the current state of our fundamental physics.

7 Failure of naturalness in cosmology and particle physics

In section 4 I focused on quantum field theories as models for condensed-matter physics, but their more famous role is as the theoretical language in which modern particle physics is written, and the idea of parameter Naturalness arose in that context. And indeed, Naturalness arguments (and the somewhat-related $O(1)$ arguments) have succeeded in explaining a number of features of the Standard Model of particle physics, and of its application in various domains. But there are two very serious exceptions, two places where Naturalness arguments seem badly to fail.¹¹

The first, in cosmology, concerns the so-called ‘cosmological constant’ (normally written Λ), originally introduced by Einstein as a parameter in classical general relativity, which describes (speaking very loosely) a uniform repulsive force permeating space. As measured in terrestrial units, the cosmological constant is extremely small: for decades it was believed to be exactly zero, but current cosmological data gives it a value (Planck Collaboration 2018) of

$$\Lambda \sim 10^{-52} \text{m}^{-2} \text{ (measured)}. \quad (4)$$

As we saw in section 2, though, it isn’t really meaningful to say of a quantity with dimensions that it is large or small. We can get a dimensionless expression for the size of the cosmological constant in classical cosmology by expressing it in terms of a cosmological lengthscale: the observed size of the Universe, $R \sim 10^{10}$ light years $\sim 10^{25}$ m. On this scale, the dimensionless value of Λ is ~ 0.01 .

Although we lack a fully consistent *quantum* version of general relativity (this is the as-yet-unsolved problem of quantum gravity), we can represent it fairly successfully as an effective field theory with a cutoff at the so-called ‘Planck scale’, or $\sim 10^{-34}$ m. Understood as a quantum field theory, the cosmological constant is a relevant parameter: its effective size on scale L ought to be \sim

¹¹There is a third exception, the so-called ‘ θ -vacuum; for reasons of space and of expository simplicity I omit it.

$(L/l_p)^2$, so that it becomes larger at large distances. On cosmological scales, that would predict a size for the cosmological constant of $\sim 10^{118}$. Or in standard units,

$$\Lambda \sim 10^{68} \text{m}^{-2} \text{ (derived from Naturalness)}. \quad (5)$$

That is: if we estimate the size of the cosmological constant using Naturalness, we are out by 120 orders of magnitude.

I should remind the reader that we cannot address this problem just by choosing an extremely small value for Λ at the cutoff scale. To obtain the observed value of Λ , we require the cutoff-scale parameter to be of $O(1)$ — but we also require it to have a very specific value, a value which must be specified to ~ 120 decimal places.

This drastic failure of Naturalness arguments has been known to, and worrying to, high-energy physicists at least since a highly-influential paper by Weinberg (1989), but has been brought into sharper focus since the discovery that the cosmological constant is not exactly zero — I think because it is much easier to imagine some mechanism that *exactly* cancels the various quantum contributions to Λ than to imagine a mechanism that just cancels them to one part in 10^{120} . (I should say that the field-theoretic arguments I discuss here are somewhat controversial, at least outside the high-energy physics community: see Bianchi and Rovelli (2010), Saunders (2002), Koberinski (2017) for some criticism. My concern in this paper is not the details of the QFT but the conceptual implications if the mainstream view is correct.)

The other failure of Naturalness is of more recent origin.¹² The mass of the Higgs boson — recently and triumphantly discovered by the Large Hadron Collider (LHC) — can be understood as the low-energy form of one of the parameters in the Standard Model of particle physics. As such, Naturalness considerations would lead us to predict that the Higgs mass is of order the Standard Model’s as-yet-unknown high-energy cutoff. The Higgs mass (in the standard units of particle physics) is $\sim 100\text{GeV}$, so that the Standard Model can be predicted to break down — and new physics to emerge — on energy scales not much higher than that. And indeed, arguments of this kind were one major reason why theoreticians expected the discovery of the Higgs to be just the overture to a new era of beyond-Standard-Model physics in the LHC.

That era has not dawned yet, and it doesn’t look as if it’s going to any time soon. No new physics has been found in the energy range that the LHC probes — which by now has reached $\sim 10,000\text{GeV}$, $100 \times$ the Higgs mass — and while it remains possible that it’s there but we aren’t looking for it in the right way, that possibility is now looking increasingly remote. For this to be consistent with our current understanding of QFT, the parameters of the Standard Model need to be at least fairly unNatural — achieving a Higgs mass only 0.01 of the cutoff energy requires the bare parameters to be tuned to one part in 10,000. This is trivial compared to the one part in 10^{120} required for the cosmological constant; on the other hand, while that argument relied on at least somewhat-speculative quantum-gravity considerations, the Higgs argument relies only on

¹²For more details see Williams (2015), Giudice (2017), and references therein.

the physics of the Standard Model, which we think we understand pretty well. What the Higgs unNaturalness argument lacks in drama it makes up for in solidity.

8 Understanding Naturalness failure

What are we to make of this dramatic failure of Naturalness in physics? One response is simply to reject Naturalness, and to claim that we had no good reason to expect it to hold in the first place. (See Hossenfelder (2018) for a clear exposition of this viewpoint.) This paper shows that the costs of doing so are much higher than are usually appreciated. In the first place, this would make the Standard Model, and general relativity, unNaturally emergent theories, the only known cases of such. In these theories as in no others, the large-scale physics is not simply extractable from the lower-scale physics through a Naturalness assumption; rather, its detailed form is encoded in the precise values of the lower-level parameters. In the case of the cosmological constant in particular, where the level of precision is one part in 10^{122} , no remotely feasible experiment or calculation will *ever* tell us that that precise value is: for ever after, one of the fundamental constants of Nature will be specifiable with sufficient accuracy *only* by saying “it is that value such that the emergent low-level physics is such-and-such”.

But the consequences are wider than this. We have seen that Naturalness assumptions are the glue that links physical explanations at different levels: if we simply reject their legitimacy then we undermine almost everything we know about inter-theoretic reduction in physics. If Naturalness arguments just brutally fail in particle physics and cosmology then there can be no reliable *methodological* argument for assuming them elsewhere (say, in statistical mechanics). And Naturalness failure in particle physics and cosmology seems to block the obvious routes by which Naturalness might have a physical justification. Yet Naturalness *does* seem to be justified, one way or another: physics (we are stipulating) is filled with examples of successful inter-level reductions which rely on Naturalness, and whose success becomes inexplicable if Naturalness after all fails. If our world were filled with unNaturally emergent phenomena then the Higgs mass and the cosmological constant would comprise just two more examples of such, but it is extremely difficult to see what coherent worldview we could create which has Natural emergences everywhere except, brutally, in these few places.

(I should note in passing that an apparent failure of the $O(1)$ assumption would have no such drastic consequences: the presence of a very large dimensionless free parameter in our most fundamental physics would be a surprising deviation from our inductive expectations, but does not seem inherently unacceptable. Much criticism of Naturalness — and in particular Hossenfelder’s view that Naturalness is just an aesthetic preference of physicists — seems to conflate the $O(1)$ hypothesis with the narrower conception of Naturalness I discuss here.)

The alternative to wholesale abandonment of Naturalness is to seek some understanding of how observed Naturalness failure could be achieved without simply baking it into an unNatural condition in our underlying physics. There is actually precedent for this elsewhere, in condensed-matter applications of QFT. Recall that in those cases relevant parameters become extremely large at large scales unless the underlying parameters are extremely finely tuned; nonetheless, there are modelling contexts where this fine-tuning is appropriate. Specifically, small values of the relevant parameters correspond in some contexts to systems on the cusp of a phase transition, such as the transition of a metal from one magnetic state to another, or of water from liquid to gas.¹³ Since systems do after all sometimes end up in phase transitions, it must be possible here to understand why Naturalness fails for the parameters, and indeed we can understand it: there is feedback between the macroscopic parameters of the system (in particular the temperature and the heat capacity) and the microscopic parameters, so that the system can evolve towards a state where the transition is occurring and the parameters are appropriately small.

Reproducing any such explanation in high-energy physics requires contingency of physical parameters (which as we have seen is speculative, albeit borne out in string theory), along with some feedback mechanism (which is even more speculative). For example, Giudice (2017) has suggested that the physics of the Standard Model might display phase-transition feedback similar to that observed in condensed-matter physics; Smolin (1997) has proposed that the values of the parameters evolve by some analog of natural selection, that drives them towards values that optimise for the creation of black holes. It is far too early to know if any such proposal will be compatible with a fuller understanding of the underlying physics (much less testable).

A rather different class of explanations starts with the observation (present in Weinberg’s original paper) that it is fortunate for us that the cosmological constant is so small: were it not, the Universe would have collapsed to nothing, or expanded to emptiness, long before intelligent life could possibly have developed. Similar arguments can be advanced for the Higgs mass, and indeed for many of the dimensionless parameters in the Standard Model. This suggests a modification to the Naturalness hypothesis:

Anthropic-conditional Naturalness: Select parameters and initial conditions by a Natural probability measure *restricted* to that part of parameter space and state space that is compatible with the subsequent appearance of intelligent life.

This is a substantive physical hypothesis, not just a methodological truism. Of course, since we emphare intelligent life, the parameters and initial conditions must be compatible with the existence of intelligent life — but equally, since the Earth is spherical, they must be compatible with a round Earth, but we do not thereby think that the shape of the Earth has been explained. As such it is

¹³In fact the *generic* liquid-gas transition is not so modelled: only the so-called ‘triple-point’ transition is. The details are not relevant here.

a hypothesis that comes in many variations: from requiring just the existence of stars and galaxies, through the existence of self-replication and Darwinian natural selection, through to the existence of intelligent tool-using creatures, all the way to the existence of beings capable of formulating the hypothesis of Anthropic-conditional Naturalness (Pratchett 1996, pp.144-5). Indeed, it is a hypothesis that makes testable predictions, albeit coarse-grained and imprecise ones: our universe will not be any *more* unnatural than is needed for intelligent life. (There have been arguments made in the physics literature, going back to Weinberg (1989) that this prediction is confirmed, at least to some degree; see Polchinski (2015) for some discussion.)

As a brute hypothesis, Anthropic-conditional Naturalness is simply a way (albeit quite an elegant way) of stating an unNaturally-emergent physics: a world conforming to it is a world where the fundamental form of the laws of physics can be given from the bottom up, but the parameters and/or initial conditions are irreducibly specified by reference to higher-level concepts like life or intelligence. But there is also a tradition of trying to explain it from something else — or rather, there are two traditions.

The first tradition requires as ingredients a mechanism for parameter contingency, a stochastic or quantum-random process that selects those contingent parameters, and a multiverse — that is, a vast number of realisations of the physics of the observed universe, each with its own values of the parameters. The proportion of worlds with given values of the parameters will be determined by some Natural measure, but ordinary observer selection means that of course our universe will be in the sub-population of worlds that are life-compatible. The second tradition has an older lineage: of course our universe would be compatible with intelligent life if it is the creation of a Designer who built it that way.

Neither explanation has much value in isolation: if the only reason to hypothesise a multiverse, or a Designer, was to explain Anthropic-conditional Naturalness, we might as well postulate the latter directly and be done.¹⁴ Things would be otherwise if we possessed independent reasons to accept either hypothesis.

At least in the case of the multiverse, we do seem to have such reasons: string theory provides a mechanism for random generation of constants (the emergence of low-energy physics from a bewildering range of possible ways spacetime can be compactified) as well as a mechanism for generating a multiverse (indeed, two: the Everettian multiverse that follows from a literal interpretation of the quantum-mechanical structure of the theory, and eternal inflation). And so (as has long been recognised; see Polchinski (2016), and references therein, for details) if string theory is correct then it provides a bottom-up justification of something like Anthropic-conditional Naturalness. The question of whether we have independent reasons to posit a Designer lies far beyond the scope of this

¹⁴Dennett (1995, ch.7) argues that the multiverse hypothesis at least offers a proof of concept that there could be *some* bottom-up explanation for the universe supporting life, which seems fair enough; but that is a motivation for trying to develop and test theories along those lines, not a reason for confidence in such a theory ahead of its development (not that Dennett says otherwise).

paper! — suffice it to say that at least within contemporary science, I see no other explanatory gap that needs filling.

For completeness I should mention one last, distinctly strange possibility. In my discussion of the explanation of Naturalness, I appealed repeatedly to the absence of “some remarkable coincidence” in the mathematics of inter-theoretic relations, in order to justify the assumption that a Natural distribution over microconditions or micro-level parameters did not give rise to a strongly unNatural distribution at a higher level. But there is admittedly something strange about talk of coincidence or improbability when we are discussing mathematics, whose truths are true *a priori*. It is not possible to be *completely* certain that such a ‘remarkable coincidence’ does not in fact occur, and if it did it would of course be true as a matter of logic. In a scenario like this, the parameters of the Standard Model and of general relativity would be mathematically determined in a unique way from some simple microphysical starting point, and yet that unique way would just so happen to fix them at exactly the right value to give rise to the unNatural physics that we observe, and in particular to support the existence of intelligent life.

I have met physicists and philosophers who say that this would remove all the mystery; I think this is exactly the wrong conclusion to draw. The possibility of high-level emergent structure being encoded directly not just into precisely-tuned values of the microphysical constants but into the very principles of mathematics would be a strangeness beyond anything we have seen before, in wild conflict with the way we normally think about physics and math. An analogy may help (adapted from Sagan (1985)): suppose that between the billionth and two-billionth digits of π we were to find, encoded in Morse code, a word-for-word copy of Newton’s *Principia Mathematica* (or, to be more provocative, of the Old Testament).¹⁵ If it were there, it would be there *a priori*, as a matter of logic — yet to react to such a discovery by shrugging and saying ‘we can prove that this had to happen, so it’s not mysterious’ would seem myopic.

9 Conclusion

Naturalness assumptions are so ubiquitous in physics, and so apparently innocuous, that they are seldom remarked upon, yet they play an indispensable role. With them, in case after case we can understand the physics of higher-level systems as emergent from lower-level descriptions, and can at least aspire to a rich and sophisticated reductionism that explains any physical system’s dynamics through chains of coarse-graining from fundamental physics. Without Naturalness, on the other hand, the connections between physics at different levels are severed and we lose any ability to understand inter-level relations.

This means that the longstanding failure of naturalness in cosmology, and

¹⁵No doubt *somewhere* in π there exist both books, and indeed this very article, encoded in Morse code or in any other coding we choose: it is, after all, infinitely long. It is finding high-level structure so relatively close to the beginning of that infinite sequence that would be incomprehensible.

the more recent evidence for Naturalness violation in particle physics, have ramifications far beyond those specific and esoteric fields. Naturalness failure here undermines arguments for Naturalness anywhere, and calls out urgently for understanding. No such explanations are trivial in their consequences: while bottom-up explanations require surprising changes in our methods of explanation, simple rejection of Naturalness as a principle involves— one way or another — a far deeper and stranger shift in our scientific world-view.

References

- Albert, D. Z. (2000). *Time and Chance*. Cambridge, MA: Harvard University Press.
- Albert, D. Z. (2015). *After Physics*. Boston: Harvard University Press.
- Anderson, P. W. (1972). More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science* 177, 393.
- Barrow, J. D. and F. J. Tipler (1986). *The Anthropic Cosmological Principle*. Oxford: Oxford University Press.
- Bassi, A. and G. Ghirardi (2003). Dynamical reduction models. *Physics Reports* 379, 257.
- Bianchi, E. and C. Rovelli (2010). Why all these prejudices against a constant? <https://arxiv.org/abs/1002.3966>.
- Bohm, D. (1952). A Suggested Interpretation of Quantum Theory in Terms of “Hidden” Variables. *Physical Review* 85, 166–193.
- Carroll, S. (2017). *The Big Picture: On The Origins of Life, Meaning and the Universe Itself*. New York: Penguin Random House.
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Oxford University Press.
- Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Chalmers, D. J. (2008). Strong and weak emergence. In P. Clayton and P. Davies (Eds.), *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*, pp. 244–256. Oxford: Oxford University Press.
- Dennett, D. C. (1995). *Darwin’s Dangerous Idea: Evolution and the Meanings of Life*. Simon and Schuster.
- Dennett, D. C. (2001). The zombic hunch: Extinction of an intuition? *Royal Institute of Philosophy Supplement* 48, 27–43. Reprinted in Daniel C. Dennett, *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*, MIT Press: Cambridge, MA, 2005.

- Dürr, D., S. Goldstein, and N. Zanghi (1996). Bohmian mechanics as the foundation of quantum mechanics. In J. T. Cushing, A. Fine, and S. Goldstein (Eds.), *Bohmian Mechanics and Quantum Theory: An Appraisal*, pp. 21–44. Dordrecht: Kluwer Academic Publishers.
- Ellis, G. F. R. (2012a). On the limits of quantum theory: contextuality and the quantum-classical cut. *Annals of Physics* 327, 1890–1932.
- Ellis, G. F. R. (2012b). Top-down causation and emergence: some comments on mechanisms. *Interface Focus* 2, 126–140.
- Frigg, R. (2009). Typicality and the approach to equilibrium in Boltzmannian statistical mechanics. *Philosophy of Science* 76, 997–1008. Available online at <http://philsci-archive.pitt.edu>.
- Ghirardi, G., A. Rimini, and T. Weber (1986). Unified Dynamics for Micro and Macro Systems. *Physical Review D* 34, 470–491.
- Giudice, G. F. (2017). The dawn of the post-naturalness era. <https://arxiv.org/abs/1710.07663>.
- Goldstein, S. (2001). Boltzmann’s approach to statistical mechanics. In J. Bricmont, D. Dürr, M. Galavotti, F. Petruccione, and N. Zanghi (Eds.), *In: Chance in Physics: Foundations and Perspectives*, Berlin, pp. 39. Springer. Available online at <http://arxiv.org/abs/cond-mat/0105242>.
- Greene, B. (1999). *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*. New York: W. W. Norton.
- Hartle, J. B. and S. W. Hawking (1983). Wave function of the universe. *Physical Review D* 28, 2960.
- Hossenfelder, S. (2018). Screams for explanation: Finetuning and naturalness in the foundations of physics. <https://arxiv.org/abs/1801.02176>.
- Jaynes, E. (1957a). Information theory and statistical mechanics. *Physical Review* 106, 620.
- Jaynes, E. (1957b). Information theory and statistical mechanics ii. *Physical Review* 108, 171.
- Koberinski, A. (2017). Problems with the cosmological constant problem. <http://philsci-archive.pitt.edu/14244/>.
- Loewer, B. (2002). Determinism and chance. *Studies in the History and Philosophy of Modern Physics* 32, 609–620.
- Maudlin, T. (2007). What could be objective about probabilities? *Studies in History and Philosophy of Modern Physics* 38, 275–291.
- Myrvold, W. (2014). Probabilities in statistical mechanics. In C. Hitchcock and A. Hajek (Eds.), *Oxford Handbook of Probability and Philosophy*. Oxford University Press.
- Planck Collaboration (2018). Planck 2018 results. VI. cosmological parameters. 78 authors. arXiv: 1807.06209.

- Polchinski, J. (1999). Effective field theory and the Fermi surface. Lectures presented at TASI 1992; arxiv:hep-th/9210046.
- Polchinski, J. (2015). String theory to the rescue. <https://arxiv.org/abs/1512.02477>.
- Polchinski, J. (2016). Why trust a theory? some further remarks (part 1). <https://arxiv.org/abs/1601.06145>.
- Pratchett, T. (1996). *Pratchett, Terry*. London: Gollancz.
- Sagan, C. (1985). *Contact*. New York: Simon and Schuster.
- Saunders, S. (2002). Is the zero-point energy real? In M. Kuhlmann, H. Lyre, and A. Wayne (Eds.), *Ontological Aspects of Quantum Field Theory*, pp. 313–344. Singapore: World Scientific.
- Smolin, L. (1997). *The Life of the Cosmos*. New York: Oxford University Press.
- Tong, D. (2009). Lectures on string theory. <https://arxiv.org/abs/0908.0333>.
- Volchan, S. B. (2007). Probability as typicality. *Studies in the History and Philosophy of Modern Physics* 38, 801–814.
- Wallace, D. (2010). The logic of the past hypothesis. Available online at <http://users.ox.ac.uk/~mert0130/papers.shtml>.
- Wallace, D. (2012). *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation*. Oxford: Oxford University Press.
- Wallace, D. (2016). Probability and irreversibility in modern statistical mechanics: Classical and quantum. To appear in D. Bedingham, O. Maroney and C. Timpson (eds.), *Quantum Foundations of Statistical Mechanics* (Oxford University Press, forthcoming).
- Wallace, D. (2017). Quantum field theory. Forthcoming in Knox and Wilson (ed.), *Handbook of the Philosophy of Physics* (Routledge).
- Weinberg, S. (1987). Newtonianism, reductionism and the art of congressional testimony. *Nature* 330, 433–437.
- Weinberg, S. (1989). The cosmological constant problem. *Reviews of Modern Physics* 61, 1–23.
- Williams, P. (2015). Naturalness, the autonomy of scales, and the 125 gev Higgs. *Studies in the History and Philosophy of Modern Physics* 51, 82–96.
- Williams, P. (2018). Two notions of naturalness. <http://philsci-archive.pitt.edu/15347/>.