

Sloppy Models, Renormalization Group Realism, and the Success of Science

July 7, 2023

Abstract

The “sloppy models” program originated in systems biology, but has seen applications across a range of fields. Sloppy models are dependent on a large number of parameters, but highly insensitive to the vast majority of parameter combinations. Sloppy models proponents claim that the program may explain the success of science. I argue that the sloppy models program can at best provide a very partial explanation. Drawing a parallel with renormalization group realism, I argue that it would only give us grounds for a minimal kind of scientific realism. Nonetheless, the program can offer certain epistemic virtues.

1 Introduction

The “sloppy models” research program originated in systems biology in the 2000s, but has seen widespread applications across a range of fields, including the study of quantum systems, neural networks, particle accelerator design, insect flight, and critical phenomena in condensed matter (?). Roughly speaking, a sloppy model is dependent on a large number of parameters, but exhibits an intriguing property: the predictions of the model are highly insensitive to the vast majority of combinations of these parameters (the “sloppy parameter combinations”), but are highly sensitive to a small number of parameters or combinations (the “stiff parameter combinations”). The proponents of the sloppy models program have suggested that it may provide an explanation for the success of science (?).

The reason is this. Many real world systems depend on vast numbers of parameters. We might think it would be almost impossible to form successful scientific theories about such systems, as they are too complex. There are far too many factors to take into consideration when forming an adequate model, one capable of making accurate predictions, useful explanations, and latching onto real regularities within the system. Nonetheless, in practice, scientists have formed highly successful effective models of many such systems. The proponents of sloppy models argue that this is because many real world systems are sloppy: in practice, they are highly insensitive to the vast majority of combinations of these parameters. Therefore, scientists can form highly successful,

simplified *effective* models of these systems, dependent on many fewer parameters, and effectively ignoring the sloppy parameter combinations.

Some the arguments put forwards by proponents of sloppy models are closely related to the ongoing philosophical debate around scientific realism. In particular, I will argue that the sloppy models program is closely related to a program in the philosophy of physics: the renormalization group realism approach. Renormalization group realists argue that renormalization techniques can provide us with grounds for a selective scientific realism about a range of effective quantum field theories.

In this paper, I will address two closely related questions. First, can sloppy models can explain the success of science? Second, if the sloppy realist explanation for the success of science is right, does it give any support for a selective scientific realism about some of our scientific models?

I argue that the sloppy models program can provide, at best, only a partial explanation of the success of science. However, it may help us to reframe the question, at least in some cases: why do so many real world systems seem to be sloppy? The proponents of the sloppy models program have not yet provided a convincing answer to this question, although there are some promising hints. Furthermore, I argue that even if the sloppy success argument were right, it would only give us grounds for a very minimal kind of scientific realism. Nonetheless, the sloppy models program brings many epistemic virtues. Sloppy models are likely to be robust against certain forms of uncertainty, about the sloppy parameters, but we cannot protect them from unconceived alternative theories. Furthermore, the sloppy models program provides a unifying account of many types of intertheoretic reduction, including some types of coarse-graining and renormalization group methods, relevant to a wide range of scientific disciplines.

In section ?? I define epistemic scientific realism, and its pessimistic challenge: this will be relevant to some of the later arguments. In section ??, I explain sloppy models in more detail, and reconstruct the “sloppy success argument”, that sloppy models can explain the success of science. In section ??, I summarize the effective quantum field theories program, and the arguments put forward by renormalization group realists. In section ??, I argue that, even if the sloppy success argument is right, it would only give us grounds for a very minimal kind of scientific realism. In section ??, I consider whether effective models of sloppy systems might provide a partial insulation against certain types of theory change. However, in section ??, I argue that the sloppy success argument can succeed only in part. I outline the epistemic virtues that I think sloppy models bring in section ??.

2 Computational Models and Scientific Realism

2.1 Epistemic scientific realism

Following ?, xviii, I define epistemic scientific realism as the belief that our best (mature and predictively successful) scientific theories are well-confirmed and approximately true of the world. The realist thinks that we can systemically and reliably infer truth from the

success of our scientific theories, whereas the antirealist does not. Loosely speaking, both have in mind the same notion of success, namely that the theory leads to (and ideally *continues* to lead to) novel predictions and explanations, experimentally confirmed or corroborated, and does not lead to too many falsified predictions.

The contemporary dialectic is centered on a debate about the history of science. Realists often argue that the best explanation for the success of our scientific theories is that these theories are (approximately, partially) true (?). Against this, antirealists have pointed to the history of science, arguing that many of our our past best scientific theories have ultimately been shown to be false (?). One variant of this is the challenge of unconceived alternatives: scientific theories are often replaced with new theories that had not even been considered before (??). If the scientific realist wishes to assert that our current best theories are approximately true, they must deny that we should expect such theories to be overturned, replaced by an unconceived alternative, as many of our past best theories were. A common realist strategy is to adopt a form of selective realism, identifying those parts of scientific theories that seem to survive theory change. Perhaps then, we have reason to believe in the truth of those parts of our best scientific theories that we can reliably anticipate will not get overturned (????).

The antirealist need not deny what ? terms the “Maddy/Wilson principle”¹. This is the claim that there is *some* systemic relationship between the description of the world offered by successful scientific theories and how things actually stand in the world. However, the antirealist must deny that we can infer the approximate *truth* of a theory (or specific parts of a theory) from this systemic connection. ? argues that selective realists must satisfy the “trust” requirement: they must account for which parts of our best scientific theories they expect to be preserved in a principled way *before* theory change in fact takes place (see also ?).

2.2 Computational models

Computational scientists in fields like systems biology use the term “model” in a more specific sense than its wider usage in philosophy of science.² Suppose that we want to learn about some target system; a model is a mathematical function that we use to generate predictions about this system. Suppose that we have experimentally extracted some finite number of measurements, which we represent by a set of real numbers $y = \{y_m\}$, indexed by points $\{m\}$. A model is a function $f : \mathcal{R}^N \rightarrow \mathcal{R}^M$ from a set of N real numbered parameters $\Theta = \{\theta_n\}$, to a set of M real numbered predictions about the system, $f(\Theta)$ (generally with $M > N$). For the current purposes, we can regard \mathcal{R}^N as the parameter space (although in practice we might be interested in a more restricted parameter space, a subset of \mathcal{R}^N). Both the predictions and measurements may have associated uncertainties. We write a cost function to measure the distance of the predictions to the empirical measurements, and try to tune the parameters (i.e. set values for each θ_n) to minimize the cost. Then we can also use our model to extract

¹This is named after Penelope Maddy and Mark Wilson– see ? and ?.

²Here we shall follow some of the notational conventions of ?.

estimations of the real parameter values. For an accurate model, whose parameter values we have successfully extracted, the m predictions $f(\Theta)$ should closely approximate the m measurements, y .³

What might our state of knowledge of the target system look like at any given time? A simplified picture might look like this. We have some preferred model, f , which best reflects our current knowledge of the target, system. However, alongside f , we likely have under consideration a wider family of (often related) models, $F = \{f, f', f'', \dots\}$. This is often referred to as the “model space” (however, note that the family may be quite arbitrary and restricted in certain cases). The parameter spaces of these models may sometimes, but will not always, coincide. Each model will have a corresponding set of best-estimates for that model’s parameters, $\Theta, \Theta', \Theta'', \dots$, with associated uncertainties. This picture is close to how some computational scientists describe their state of knowledge (for instance see ?????). However, there is great flexibility in precisely how we define the models and parameters, and specify our model and parameter spaces. For instance, we might choose to reduce the size of the model space under consideration by including additional parameters and suitably generalizing the models.

For example, let us consider, the widely used ? epidermal growth factor receptor (EGFR) model of biological signaling (see also ?????). The model f describes the dynamics of several biochemical reactions, using a system of 15 independent differential equations, with a space of 48 parameters. For example, many of these parameters tune the various reaction rates. However, this is not the only model considered for this system. One alternative “mechanistic model”, f' , (see ?) adds several additional reaction steps, involving new chemical species, with a larger space of 70 parameters. Observe that in this particular case, f and f' are closely related but use a different parameter space (\mathcal{R}^{48} and \mathcal{R}^{70} respectively).

2.3 Realism about computational models

Let us try to explicate the realism-antirealism debate in the context of computational models. Realists believe that our best models probably represent the target system accurately, by latching onto genuine regularities within the target system, rather than merely matching the data (such models might nonetheless be approximations of the actual target system and need not be unique). I will call such models “latching models” of the target system. These regularities could be encoded in different ways in the model, through the functional form of the model, some of the parameter values, or both. Not only should some of the predictions of a latching model accurately match the measurements $y_m \in \mathcal{R}^m$; the model should be able to continue generating at least some further correct predictions. In fact, the realist makes stronger demands: we should be able to anticipate some predictions that the model will continue to get right in advance. Against this, anti-realists argue that we cannot *know* that our best models are latching, (and we

³These kinds of models are described as “computational” as they are usually sufficiently complicated that we must use computer simulations to extract numerical predictions. Whilst occasionally the models are simpler than this, the term is now widely applied to this general approach (?).

may have positive some reasons to doubt that they are), or at least cannot know which correct predictions it will continue to make.

Consider the sort of typical state of knowledge described in section ???. We need to consider three ways our model might change if our knowledge of the target system improves:

1. **Change within parameter space:** we keep the same model, but our best estimates of the parameters take different values: $\Theta \rightarrow \Theta_{\text{new}}$. Here, $\Theta, \Theta_{\text{new}} \in \mathcal{R}^N$; the parameter space has not changed.
2. **Change within a well-defined model space:** we replace our theory with a new one using a different model within the set of models that we are considering: $f(\Theta) \rightarrow f'(\Theta')$. Here, $f, f' \in F$, with $\Theta \in \mathcal{R}^N$ and $\Theta' \in \mathcal{R}^{N'}$. In general, the parameter spaces may have changed.
3. **Change to an unconceived alternative:** we replace our model with a new one, not even represented in the previously considered model space: $f(\Theta) \rightarrow g(\Theta'')$. Here, $g \notin F$, and again the parameter spaces may differ.

Consider again the ? EGFR model, and imagine that we take new measurements of the target system. Suppose that, when we use the new measurements to extract best estimates of the reaction rates, we find that the values have changed. This would be an example change within the parameter space. Alternatively, suppose that we find that the model f struggles to match our experimental results, no matter how much we tweak the parameters. However, the mechanistic model, f' seems to perform much better, overall. We decide to adopt the mechanistic model as our preferred model of the target system. This would be an example of a change within the model space. Finally, suppose that neither model, nor any of the other standard models, seem to match our experimental results. Eventually, we are force to shift to a radically different model that we had not even considered previously. This would be an example of a shift to an unconceived alternative.

Change within parameter space is not usually a major threat to the scientific realist: the model remains the same, although tuning parameters may involve changes in the relative importance we give to various processes and their interactions. In principle this *could* involve substantial change: a model parameter might be used to switch on or off terms that represent entirely new processes within the target system. However, in practice, particular models usually codify a more specific set of assumptions about the target system. The potential for change within a well-defined model space may or may not be a threat to the scientific realist, depending on the set of models under consideration. A selective realist is happy to accept our best models might be replaced with better models, as long as those preserve the correctly latching regularities of our current best models. However, in general, the set of theories under consideration may not preserve the same regularities, and we may not know which regularities correctly latch onto the same processes in the target system. Finally, change to an unconceived alternative is usually a threat: an unconceived alternative could in principle completely

replace our current best model. As such, the recent debate around scientific realism has focused on this latter type of change.

Let us suppose that we seek to model some physical target system. In general, we will have some prior knowledge of the system in question. For example, we may have a list of possible features of the system that it may be necessary to represent in a model, even if we do not know the right way to do so. However, systems with many parameters will generally be harder to model than systems with fewer. Suppose that we can somehow highly restrict the features possible in a model of the target system (for example, suppose we have some reason to believe that the only possible models of the system must be polynomial equations up to some n th order). Then there may be more possible models for systems with many parameters than a system with fewer parameters (for example, there are vastly more possible polynomial equations of n th order with 1000 parameters than there are such equations with only 10 parameters).⁴

This leads to a puzzle with regards to the success of the computational sciences. The systems studied in systems biology, climate science, condensed matter physics and other largely computational fields are highly complex, relying on enormous numbers degrees of freedom, few of which are precisely measured or well-understood. For example, the internal workings of a biological cell will in general rely on the interactions between many thousands of different proteins. A priori, one might expect that the task of successfully modeling such systems would be almost impossible: there are too many different factors affecting the system to adequately account for all of them and their interactions⁵. Nonetheless, researchers frequently produce predictively successful effective models of complex systems such as these. The success of these effective models, in generating predictions for systems far more complex than the simple model, demands an explanation. Sloppy models proponents claim that they can offer an explanation that solves this puzzle.

3 Sloppy models

There is no single, strict definition of a “sloppy model”. However one might differentiate two main approaches, which I will call the coordinate-dependent and the coordinate-independent definition (see ?). We will need to discuss both.

3.1 The coordinate-dependent definition

In the loosest sense, a model is “sloppy” if it depends on a large number of parameters, but is highly insensitive to the vast majority of combinations of these parameters (the “sloppy parameter combinations”). We can greatly change the values of these sloppy

⁴The problem of finding the right model of a system in this way can be thought of as special case of the familiar problem of induction, as applied to the task of statistical modeling (for example, see ?.)

⁵This general problem is particularly relevant in the computational sciences. In linearized statistics, the task of fitting a model to noisy data can become harder when we include more parameters in the model. Eventually, the parameters will contribute towards tuning the model to the noise, rather than the signal.

parameter combinations, perhaps by factors of thousands or tens of thousands, without significantly changing the predictions generated by the model. On the other hand, the model might be highly sensitive to a small number of parameters or combinations (the “stiff parameter combinations”). Another way to express this is that the $N \times M$ dimensional model might have a much lower *effective* dimensionality. As a result, it is hard to extract good experimental estimates of the model’s parameters: all but a few parameter combinations are only very weakly constrained by the model fit to the data.

We can use Fisher Information Matrix (FIM) (see appendix ?? for details) to precisify the definition. The FIM provides a measure of the information that the observed data provides about each parameter in a model. More precisely, it tells us about the expected curvature of the log-likelihood function of the observed data with respect to the model parameters.

We call the eigenvectors of the FIM the local (or sometimes “renormalized”) eigenparameters. These directions correspond to linear combinations of the original (“bare”) parameters. The eigenvectors give the directions of principal curvature, with the corresponding eigenvalues giving the corresponding magnitudes of the curvature. The curvature tells us how quickly the log-likelihood function changes in each direction. Small eigenvalues correspond to bare parameter combinations whose values have only a small effect on the model predictions, whilst large eigenvalues correspond to combinations that have a large effect. Then, sloppy models are characterized by an enormous range of eigenvalues, with an approximately exponential distribution, such that there are a small number of stiff eigenparameters and a large number of sloppy eigenparameters (see figure ??).

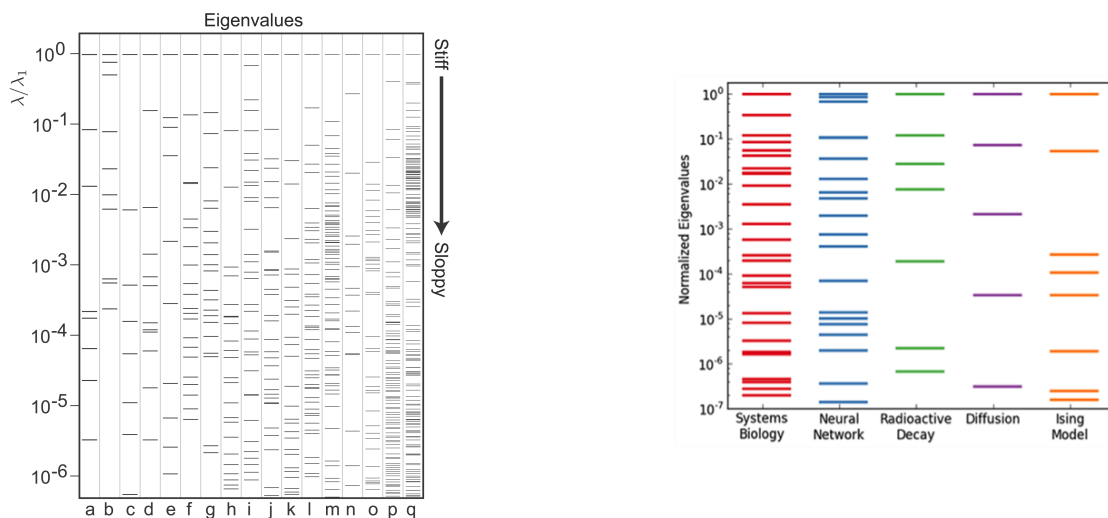


Figure 1: Left: eigenvalue spectra for 14 different systems biology models, collected by ?. . Right: eigenvalue spectra for models in a number of different fields (?????), collected by ?. In both cases, observe that the eigenvalues are approximately uniformly distributed across a log scale, and vary across many orders of magnitude.

Early on, researchers suspected that sloppiness was an artifact of the choice of model parametrization (for example, see ?). For example, we can transform a model with many sloppy eigenparameters to a model without sloppy eigenparameters by means of a simple coordinate transformation. As a result, researchers have searched for a way to define sloppiness independently of the parametrization.

Let us consider an illustrative example. Suppose that, for some system, we need to fit observed data $y_m \in [0, 1], \forall m$, and our model of the system is an N -th order polynomial, viewed as a sum of monomials, $f(y, \theta) = \sum_{i=0}^N \theta_i y^i$. The model might well be sloppy, for the reason that the monomials all have a similar same shape, being flat and near zero, and then rising quickly towards one. As such, they can easily be exchanged for each other. But we could write a model that would generate identical predictions, for example $f(y, \theta') = \sum_{i=0}^N \theta'_i L_i(y)$, where each L_i is the i th Legendre polynomial. The Legendre polynomials are orthonormal in the L^2 norm on the interval $[0, 1]$: transforming the model to this basis can completely remove the sloppiness (??).

3.2 The coordinate-independent definition

However, we can use information geometry to define sloppiness independently of the choice of parametrization. In the information-geometry approach (??), we interpret a model with N parameters and making M predictions (with the requirement $M > N$), as an N -dimensional model-manifold, with θ giving the co-ordinates, embedded in the data space \mathcal{R}^M . The parameters, θ , then serve as manifold coordinates. The observed data is a single point within the data-space, possibly lying outside of the model manifold. The FIM is symmetric and positive semi-definite: we can interpret it as a Riemannian metric on the model manifold⁶. The FIM metric measures the distance in parameter space in units of the standard deviations of the parameters, given their probability distributions $p(x|\theta)$. Then, we can interpret distances between points as the distinguishability of the predictions from different parameter choices. Now the task of fitting the model to data can be interpreted geometrically, as the task of finding the point on the model manifold that is closest to the data-point.

The model manifolds have “boundaries”: we can often take parameters or parameter combinations to their extreme values without generating infinite predictions. We can understand this better by looking at the lengths of the geodesics of the manifold, using dimensionless model parameters. This coordinate-independent approach is generally interpreted as evidence that sloppiness is an intrinsic property of the model (see ???). Whilst we can change the eigenvalues by transforming the model parametrization, we cannot change the geodesic lengths (see appendix ??)⁷. In a sloppy model, the lengths of the geodesics generally have an approximately exponential distribution, closely related

⁶For consistency with the literature, I will continue to refer to this as the “Fisher Information *Matrix*”. However, strictly speaking, when we interpret the FIM as a metric, it is not a matrix at all, but a tensor of type (0,2).

⁷Recall, however, that the geodesic lengths are measured in units of the standard deviation of the observed data. As such, there is a sense in which the sloppiness of a model is still a result of our degree of knowledge about the system.

to the distribution of FIM eigenvalues. There are some parameter combinations that can be varied across their entire range of possible values without significantly changing the model predictions. The shape of the manifold-FIM metric pair is described geometrically as a hyperribbon, an N -dimensional generalization of a ribbon, with many short dimensions and only a few longer dimensions (see figure ??).

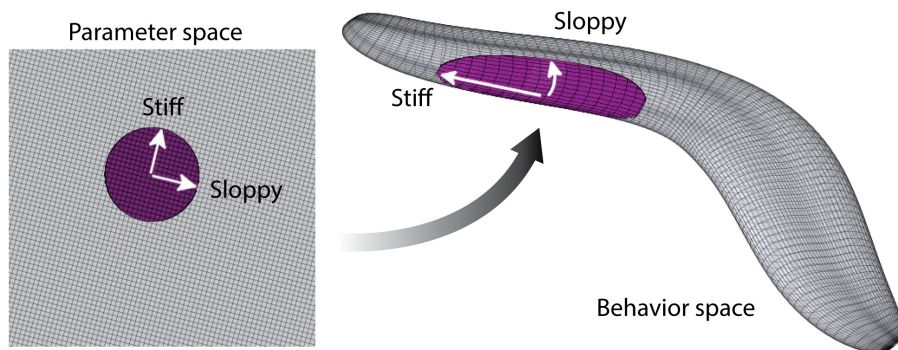


Figure 2: Schematic of a parameter space and data manifold (behavior space) for an imagined sloppy model with two parameters. Using the FIM-metric, the model has a longer (stiff) and shorter (sloppy) direction (figure from ?).

This approach endows the model manifold with enough structure to generate a family of reductions. We can generate simplified models, which depend on a smaller set of parameters, to describe the system in various limiting cases by taking particular parameters to extremal values.

More precisely, the manifold boundary approximation method (MBAM) works by iteratively applying the following procedure. First, we identify the sloppiest parameter combination (corresponding to the eigendirection with the smallest corresponding eigenvalue). Second, we construct a geodesic with the best estimates of the parameter values and this eigendirection as its initial conditions.⁸ We follow the geodesic path until a boundary is reached. Boundaries correspond to points on the manifold where the metric (the FIM) is singular. Typically this is when the parameters reach their extremal values (often specified to be 0 or $\pm\infty$; in general we will reparameterize the original model such that participating parameters are grouped together into a single combination that goes to zero at the boundary.), where the model becomes unresponsive to changes in the parameters (see ? for further details). By choosing a geodesic along sloppiest directions, we should find a nearby boundary point with nearly identical predictions for the model's parameters. Third, we evaluate the limit associated with the boundary to produce a new model with one less parameter. The reduced model represented by the submanifold is identified as an approximation of the original model under some limiting conditions. Fourth, we recalibrate the new model by fitting it to the behavior of the original model.

⁸Note that the eigendirection actually defines two directions: see ?, supplemental materials for further discussion.

This procedure can be reapplied to successfully remove the sloppiest parameters.⁹

This reduced model is often described as an “effective” model¹⁰. At least in some circumstances, the MBAM reduction method is a generalization of some well-known reduction techniques, such as taking singular limits, scale separation, equilibrium approximations, and renormalization group transformations (???)

In general, there might be many possible effective models, which we could arrive at under different MBAM dynamics. Anticipating section ??, we might think of MBAM dynamics as a flow, taking points in the manifold towards one of the model boundaries. Then we can describe the set of points in the model manifold as forming a basin of attraction of each effective model, under MBAM dynamics.

3.3 The sloppy success argument

As I explained in section ??, the success of science presents a puzzle, especially in complex fields, where systems are likely to depend on very many parameters. Advocates of the sloppy models program suggest that sloppy models provide an explanation for this success (?). Let us call their argument the “Sloppy success argument”. The key is that sloppy systems seem to be ubiquitous across these scientific fields, and these models are highly sensitive to only a small number of stiff parameter combinations. Let us call a system a “sloppy system” if the laws, rules or regularities governing the system can be well-represented using a *latching* sloppy model. Then the sloppy system can also be well-represented by at least one effective model, dependent on far fewer, stiff parameters. Such an effective model will generally be far easier to build than a more fine-grained model of a complex system that accounts for more parameters. Another way to say this is that a latching model of a sloppy system forms a basin of attraction of some effective model under a family of intertheoretic reductions, given by some MBAM dynamics.¹¹

⁹One might ask, can we not simply reduce the model to remove the sloppiest parameters, without first approaching a model boundary? In principle, this is sometimes possible. However, the sloppiest parameter combinations may be highly nontrivial combinations of nearly all the individual parameters. In general, these parameter combinations can be difficult to remove from the model. However, at the manifold boundary, the smallest eigenvalue (corresponding to the sloppiest parameter combination) will approach zero. An alternative way to visualize this is that, although the geodesic’s initial direction may have involved a complicated combination of parameters, as it approaches a boundary it rotates into a simple, physically relevant combination. As such, the geodesic paths in parameter space tend to straighten out. Furthermore, the smallest eigenvalue of the Fisher Information matrix tends to approach zero. For further details, see ??.

¹⁰We might say that such effective models are adequate for particular purposes (see ?). For example, the models are highly adequate for the purpose of constraining certain parameter combinations, but not others. Likewise, they generate good predictions for the target system under certain limiting conditions, but not in others.

¹¹Of course, some sloppy systems are quite simple, and do not *require* effective model in order to be well-understood. However, in principle at least, such systems are susceptible to the MBAM approach, so they *can* also be well-represented by an effective model, even when one is not needed (for some examples, see ?, supplementary materials). Furthermore, note that, by definition, a sloppy system can be represented by a latching sloppy model. But at least in principle, a sloppy model could still fail to be latching, for example if we are wrong about the stiff parameter combinations.

The sloppy success argument then takes the following form. Whilst it would be hard to form a predictively successful complete model of most systems, it is generally comparatively easy to form an effective model dependent on just a few parameter combinations. We assume that many such complex, real-world systems are sloppy, i.e. they have latching sloppy models. We can think of a sloppy model describing some system as forming an attractor basin of a simpler effective model under some particular MBAM dynamics. We can generate predictively successful effective models to represent limiting cases of sloppy models. Therefore, under appropriate conditions, we can form predictively successful effective models for many real world systems.

Proponents of this argument present it in both a local and global form.¹² The local argument is applied to specific systems, whereas the global argument is applied to science, or the computational sciences, in general. According to the global argument, we have good reason to believe that many real world systems are sloppy. Sloppy systems are the type of system about which we can form effective models. So we should expect to be able to form successful, effective scientific models of many real world target systems. According to the local argument, if we have a good reason to expect a particular system is sloppy, then we should expect to be able to form an effective model of this system. Therefore we should expect scientific efforts in modeling this system to be successful. However, the global and local forms are closely connected. In general, it is hard to know a priori that a real world system is sloppy. One reason to expect any particular system to be sloppy might be if we have good reason to expect that *many* such systems are sloppy.

The success of a sloppy model will be robust against certain kinds of change in our knowledge of the target system. If we learn that our previous understanding or measurements of the sloppy parameters was wrong, this is unlikely to strongly affect how reliable we should consider the effective model to be. As such, we can build predictively successful effective models of many systems about which we are highly ignorant, insofar as that ignorance pertains to the sloppy parameters.

4 Renormalization group realism

At first glance, this closely parallels a common claim in physics and philosophy, that the “effective field theories” program can explain success of our best theories in high energy and condensed matter physics, in spite of our ignorance about the underlying physical theories (?). A number of philosophers (?????) have recently marshaled the effective field theory program in the defense of a local form of scientific realism, variously labelled “effective realism” or “renormalization group realism”. This defense has been challenged by ? among others. The arguments around renormalization group realism mirror those of the general realism-antirealism dialectic, described in section ??. I will now summarize the program in more detail.

¹²See ?? for examples of the local form of the argument. See ?? for presentations of the global form of the argument.

4.1 Effective field theories

Our most successful theories describing a wide range of physical phenomena are quantum field theories (QFTs); the Standard Model of particle physics is a quantum field theory, for example. However, the QFT program was beset by problems from the outset, in particular the appearance of troublesome infinities in certain equations for interacting theories. The efforts to tame these infinities eventually led to the renormalization group approach pioneered in the 1960s and 1970s (?). I summarize some features of this approach here, necessarily simplifying considerably and eliding over most technical details (for an overview of the approach written for physicists, see ? or ?; for an introduction written for philosophers see ? and ?).

A quantum field theory is characterized by a Lagrangian, \mathcal{L} , with free parameters representing masses and charges, so called “coupling constants”, whose values tell us the strength of different types of interaction. We can derive the equations of motion from the Lagrangian, along with other physically measurable properties such as cross-sections that tell us about the rates with which different physical processes take place. To take a toy example, the so-called ϕ^4 theory Lagrangian may be written

$$\mathcal{L} = \frac{1}{2}\partial\phi^2 - \frac{1}{2}\alpha_0^0\phi^2 - \alpha_1^0\frac{\phi^4}{4!}, \quad (1)$$

where ϕ is a scalar field, and ∂ is an appropriately defined derivative operator.

However, the value of the theory’s “bare” coupling constants cannot be directly measured; rather, we measure the value of physical quantities that depend on the interaction energy at which we probe the system. We will call these quantities the physical coupling constants, α^p . In general they will be a function of the interaction energy, μ , the bare coupling constant, α_0 and a scale, Λ , that we will discuss imminently; thus we write $\alpha^p(\mu, \alpha_i^0, \Lambda)$.

Typically, attempting to obtain predicted values for physical observables results in a power series expansion, in which certain terms take the form $\int_0^\infty \frac{k^a}{k^2+m^2} dk$, where k represents the possible momenta of the interaction, and a a constant. For $a > 2$, this integral will become infinite. In order to begin taming this infinity, we “regularize” the theory. One simple way to do this ¹³ is by imposing a cutoff scale, Λ , at an energy much higher than the interactions we care about. In effect, we chop off our integral’s upper limit, yielding $\int_0^\Lambda \frac{k^a}{k^2+m^2} dk$, ignoring any contributions of higher energy and rendering our integral finite by brute force. Without any principled justification, this approach would seem to be worryingly ad hoc.

The process of rescuing our theory is called “renormalization”. We rewrite our theory in order to remove dependence on the arbitrary regularization scale, Λ . We imagine our theory as sitting in an N -dimensional parameter space, defined by the N physical coupling constants, α_i^p , $i \in \mathbb{N}$. We shift our physical coupling constants as a function of the interaction energy scale, μ , in order to absorb or cancel any dependency on the cutoff

¹³Typically, other regularization techniques are used in order to preserve the symmetries of the original theory. Regardless, the regularization procedure tames the infinities by some method, whilst introducing a regularization scale, Λ .

scale, Λ , generating “renormalized” coupling constants. In effect, we have a Lagrangian that describes the physics at every scale by shifting the value of the physical coupling constants in order to cancel out contributions from the cutoff. We call these the “running coupling constants”, and describe the process as “renormalization group flow”¹⁴.

The renormalization group flow may contain certain fixed points or fixed surfaces (the latter spanned by renormalizable couplings at low energies). At these fixed points or surfaces the value of the physical coupling constants remains unchanged as the energy scale, μ , changes. This phenomenon is referred to as “scale invariance”. A large family of theories, those with the same fields and symmetries (but potentially differing in the number and strength of nonrenormalizable interaction terms) may flow towards the renormalizable subspace, forming its “basin of attraction”.¹⁵ This property is called “universality”, although this description is misleading: in general not all high-energy theories will flow to this fixed point¹⁶. Physically, this means that the low energy physics described by this theory will be approximately invariant, regardless of the high energy behavior of the theory¹⁷. In effect, the renormalizable part of the theory can be approximately decoupled from the physics energy scales.

With this in place, we can understand many low energy quantum field theories as “effective field theories”, defined only over certain energy scales, but breaking down at other scales. We can think of the method as a kind of coarse-graining, shielding the effective field theories from our ignorance of physics at unexplored scales. Regardless of the the high energy physical processes that take place, of which we are generally highly ignorant, we arrive back at the same effective theory to describe physics at more familiar scales.

The effective field theory program has achieved remarkable success in many area of physics. Examples of effective field theories include the theory of quantum hadrodynamics, which describes for the binding together of atomic nuclei, the Fermi theory of the weak interaction, and the BCS theory of superconductivity. There are good reasons to believe that the Standard Model of particle physics is an effective field theory of this type.

4.2 Explaining the success of quantum field theories?

Physicists have constructed many predictively successful field theories to describe systems in particular regimes. However, in general, we are highly ignorant about fundamen-

¹⁴Unsurprisingly, sometimes this procedure will not be possible. A theory for which a bare coupling constant has dimensions of length ^{D} , for $D \leq 0$ will not be renormalizable. Originally, renormalizability was simply taken as a requirement for any candidate physical theory. But the modern approach explains renormalizability as a widespread feature we should expect of many physical theories. More precisely, it explains that non-renormalizable terms will become negligible as we approach experimentally accessible energy and distance scales.

¹⁵See ? for a proof of this in the case of ϕ^4 theory.

¹⁶More precisely, the theories in the universality class collapse to a finite dimensional surface of attraction at low energies (see ?, pages 652-660).

¹⁷The renormalization group method may also explain universality across certain theories that lie not *in*, but close to, the basin of attraction – see ?.

tal physics. For example, the Standard Model of particle physics has been remarkably successful at generating qualitatively and quantitatively novel predictions about the electroweak and strong forces; however, physicists have good reasons to believe that the Standard Model is not a fundamental theory. A priori, one might expect that the task of generating successful models of such systems would be almost impossible without knowledge of the underlying physics. The success of these effective field theories demands an explanation.

Advocates of the renormalization group realism program suggest that effective field theories provide an explanation for this success (??????). Let us call this argument the “renormalization group realism” argument. The key is that our best theories of physics, such as the Standard Model, should be understood as merely effective. These effective theories are a coarse-grained description of an underlying theory. Any new theory within the basin of attraction will flow close to the effective field theory. There is a sense in which the renormalization group flow can be thought of as a family of reductions, in which our effective theory can be understood as the low energy approximation of the higher energy theories within this basin of attraction (?). We can formulate this theory, even without a knowledge of the renormalized parameters. As such, we can build predictively successful effective field theories, in spite of being highly ignorant about the underlying physics.

Furthermore, the effective field theories will be robust against certain kinds of change in our knowledge of the underlying physics. Our measurements provide us the values of the physical “renormalized” coupling constants, but the values of the underlying “bare” coupling constants are insensitive to this. We are able to build predictively successful effective field theories describing many systems for which we do not have any experimental access to the bare coupling constants.¹⁸

The renormalization group realist argument then takes the following form. Whilst it would be almost impossible to generate a physical theory that accurately describes a field theoretic system at all energy scales, it is comparatively easy to form an effective field theory, accurate at some energy scale. We assume that a higher level field theory sits in (or perhaps close to) the basin of attraction of the effective field theory under renormalization group flow. Thus the effective field theory provides a good effective description of the underlying physics, when described at an appropriate scale.

In what sense is this a *realist* explanation? Renormalization group realists are selective: they hope to carve out those parts of a theory worthy of a realist commitment from those that are not, before theory change takes place. The renormalization group realist thereby hopes to respond to the challenge posed by Stanford’s trust argument (see section ??) . Insofar as we expect the true theory to be in the basin of attraction of the effective field theory, we expect theory change to be to other theories that flow to the effective field theory. Then, in principle, we can anticipate which parts of the theory will be invariant under theory change. These will be the features of the effective field theory that are invariant under renormalization group flow, and so insensitive to the details of physics at untested energy scales. Renormalization group realists suggest that these features of the theory are worthy of a realist commitment.

¹⁸However, see ? for a different perspective on the significance of the bare parameters.

However, It is not immediately obvious what this could in fact commit us to. Beyond merely the observables predicted by the theory, ? suggests correlation functions (expectation values of products of field operators associated with spacelike separated spacetime regions $x_i \dots x_n$, of the form $\langle \phi(x_i) \dots \phi(x_n) \rangle$) could be the locus for a realist commitment. The correlation functions are highly insensitive to the details of physics of high energy scales and are used to derive the measurable observables such as cross-sections. The renormalization group flow is defined so as to keep the correlation functions invariant. Nonetheless, it is unclear what precisely a realist commitment to correlation functions might entail (see ?????? for an ongoing debate).

4.3 Renormalization and Sloppiness

Hopefully, the parallels between the sloppy models program and the renormalization group realism program are already obvious. Both try to explain the success of science in some domain, about which we believe we are highly ignorant about many of the parameters or parameter combinations. Nonetheless, scientists manage to produce predictively successful, highly coarse-grained *effective* theories, or models, about the system, dependent on smaller numbers of parameter combinations. However, we expect a wide range of theories, or models, to flow to the same effective theory under some family of reductions. If the true state of the target system can be represented by some theory, or model, within that space, then it can also be well-described, in an appropriate regime, by the effective theory, or model. Therefore, we are able to construct an effective theory, or model, that is predictively successful in some appropriate domain, in spite of our ignorance of the target system.

As I already hinted, the similarity between these approaches may be more than a mere coincidence. At least some effective field theories, such as the Ising Model of ferromagnetism, are sloppy models, and there is good reason to suspect that this generalizes to other effective field theories (??). Furthermore, renormalization group flow can be thought of as a special case of the MBAM reduction process (?). Therefore, it makes sense to consider the argument that sloppy models explain the success of science alongside the existing philosophical debate around RG realism.¹⁹

5 Sloppy realism about what?

We need to distinguish two issues. First, is the sloppy success argument right, i.e. do sloppy models explain the success of science? Second, to what extent does this give us reason to be realists about all or parts of successful effective models? Let us start

¹⁹Note, however, that in both fields, work can go in both directions. In quantum field theory, researchers have usually begin with a low energy effective field theory, before arriving at the higher energy theory. However, sometimes researchers have begun with the higher level theory and then inferred the lower-energy effective theory. Likewise, in the sloppy models program, in some cases, researchers have begun with an effective model, and then moved to describe a less coarse-grained sloppy model. In other cases, they began by formulating a sloppy model, from which effective models were derived through MBAM dynamics. However, the details of this do not change the general structure of the argument.

with the second question. Assuming, for the moment, that we think the sloppy success argument gives us a good explanation of the success of certain effective models, does it also give us good reason to be realists about certain features of these models?

The advocates of the sloppy models program have not used the language of “scientific realism”. However, this may reflect a difference in disciplinary boundaries: advocates of the sloppy models program are practicing scientists, not philosophers of science. Yet, given the close parallel between the sloppy success argument, and the renormalization group realism argument, it is worth considering whether the sloppy success argument should also give us grounds for being scientific realists.

Suppose that we have some effective model about a target system, an approximation of a latching sloppy model of the system (see section ??). This latching model forms a basin of attraction of our effective model, under MBAM dynamics. As with the case of the renormalization group methods, the MBAM approach could potentially be viewed as a family of Nagelian reductions²⁰. For instance, we can derive the effective theory from any of the more general sloppy theories in its basin of attraction using bridging laws in which we take the limiting parameter combinations to extremal values. Likewise, we can infer the renormalized parameter combinations using the parameters of the more general theory.

However, this does not yet preclude an anti-realist stance. As I explained in section ??, many antirealists accept the Maddy-Wilson principle, that our scientific theories may latch onto real world regularities. Only if we can identify features of our effective model that will be invariant with respect to theory change (i.e. invariant under MBAM dynamics), can we know which features of our model might be appropriate for a realist commitment.

Can we determine those features of an effective model that are invariant under MBAM dynamics in advance? One obvious candidate could be the values of the *stiff* local parameter combinations. These do provide content about parameter combinations which we cannot be measured directly. More generally, relations between these parameter combinations may also be approximately invariant under MBAM dynamics. However, in general an interpretation of these parameter combinations will depend on their role within the model, including its non-invariant features. These non-invariant features may be replaced by theory change: only the invariant parameter combinations remain unchanged. The meaning of those invariant parameter combinations could change significantly depending on the rest of the underlying model. This would leave some, albeit fairly minimal, features of the effective model potentially suitable for a realist commitment.²¹

²⁰According to the Nagelian account of intertheoretic reduction (?), a theory T_P may be reduced to another theory T_F if two requirements are satisfied. Connectability is the requirement that, for every theoretical term in T_P , there is a corresponding term in T_F . Derivability is the requirement that the laws of T_P can be deduced from the laws of T_F , alongside some auxiliary assumptions (usually taking the form of specific boundary conditions or further idealizations). If both of these conditions are satisfied, then T_P , can be reduced to T_F . The Generalised Nagel-Schaffner Model of Reduction (??) loosens these requirements, to allow for an approximate derivability relationship between theories.

²¹Some scientific realists have argued for an extremely minimal scientific realism (?).

Let us return to the ? EGFR model, containing 15 independent differential equations, with 48 parameters. Most of the parameter combinations are sloppy. By applying MBAM, we can create an effective model, consisting of just 6 independent differential equations, dependent on 12 parameters – this model is still adequate in its intended domain, whilst carrying the advantage of being substantially simpler, conceptually and computationally. These equations generally describe limits in which the biochemical reactions of the EGFR reactions equilibrate, turn off, saturate, or never saturate. To take one example, the EGFR model describes a chain of 4 reactions involving several proteins, $C3G \rightarrow Rap1 \rightarrow BRaf \rightarrow Mek \rightarrow Erk$ ²² specified by eight parameters, whereas the effective model describes a direct interaction, $C3G \rightarrow Erk$, specified by just a single parameter, ϕ . However, ϕ can be written as a limiting approximation of a nonlinear combination of these eight parameters. The effective model summarizes the chain of reactions and allows us to generate accurate predictions in an appropriate limit. As the authors note,

The simplified model, therefore, contains real biological insights and . . . serves as the basis for understanding and predicting the functional effects of microscopic perturbations, such as mutations or drug therapies, on the system’s macroscopic behavior. (? , page 3)

In this case, it seems the relation between C3G and EGFR latches onto a regularity in the more complex model. This relation might then be a suitable target for a realist commitment. However, remember that the local parameter combinations are generally not the same as those natural parameters that we measure. If the local parameter combinations are very unnatural, it may not be obvious what a realist interpretation of them should entail.

Of course, as with any formalism, the stiff parameter combinations, and relations between those parameters, i.e. those parts of the effective model that might be suitable for a realist commitment, only carry a meaningful interpretation within a specified model. The interpretation of the invariant parameter combinations may be model-dependent, even though the parameter values themselves remain unchanged. In the EGFR example, the parameter ϕ appears to codify a single reaction rate in the effective model, whereas in the full model it codifies a more complex relation, involving several different reaction rates.

Nonetheless, if the sloppy success argument is right, it would seem to give us grounds for a minimal realism about some features of effective models. Overall, however, this is not a very surprising conclusion, given the strength of the assumptions we have made. In particular, this has all rested on a key supposition– that we *know* that the target system can be represented by a latching model, within the attractor basin of an effective theory under MBAM dynamics.

²²More fully, these proteins are Crk SH3-domain-binding guanine nucleotide-releasing factor, Ras-proximate-1, B-Rapidly Accelerated Fibrosarcoma, Mitogen-activated protein kinase kinase and Extra-cellular signal-regulated kinase – see ?? for a more detailed overview and discussion of the proteins and their role in the EGFR model.

6 What types of success can sloppy models explain?

Proponents of the sloppy success program believe that sloppy models can solve the puzzle presented in section ??, and in so doing explain the success of science. We now have the tools in place to analyze this claim a little more precisely. Let us first assess *what* sorts of success sloppy models might have the potential to explain.

If we have a sloppy target system, then we may be able to form an effective model of that system, one that is simpler and dependent on fewer parameters than a more complete model of the system. If the sloppy success argument can explain the success of science, then these effective models need to be latching models of the target system. Then, we should expect that they will be insulated against certain types of theory change. If the effective model is a latching model, then we might expect theory change to take us only to other latching models within the basin of attraction of the effective model, under MBAM dynamics. In other words, the effective model will only be replaced by other models, of which it is a reduced version.

Unfortunately, in general we do not know whether any particular model that we devise really is a latching model of the system. A model that we think is latching could merely be predictively successful under the conditions and scales that we have currently tested. Perhaps the model would eventually be replaced by an altogether different theory. In the extreme case, the target system could be better described by an unconceived alternative—some model that we have not yet considered. Recall, the three types of theory change discussed in section ??: change within a parameter space, change within a well-defined model space and change to an unconceived alternative. Effective models of sloppy systems may be insulated against certain kinds of theory change, but will not be insulated against others.

First, change in parameter space is unlikely to be a problem for an effective model. By the definition of a model in section ??, a single model can accommodate a range of parameter values (and even the predictions of the model are generally robust against changes in the sloppy parameter values). The model is flexible enough to accommodate this sort of change in our knowledge of the target system. Second, an effective model may be robust against change within a well-defined space of models, if those models are in the basin of attraction of the effective model. Insofar as we only consider theory change within a family of models related by MBAM dynamics, then the effective model will only change to models of which it is a reduced version. However, an effective model does not provide any way to guard against change to a fundamentally different theory in model space, or to an unconceived alternative outside of model space. Unfortunately, as I explained above, it is precisely this sort of theory change that forms the crux of the contemporary philosophical debate.

The parallel with the renormalization group realist program is instructive here. ? cites the effective field theory methodology as an argument against unconceived alternatives—after all, it gives us a means to anticipate at what scales a theory might break down, and what theory might replace it (see ?). For example, it was predicted long in advance that one effective theory, the Fermi theory of the weak interaction, would

start to fail at around the 100 GeV scale, long before it was replaced by a generalized electroweak theory (see ?, page 5).

However, the renormalization group program can only help us to anticipate theory change within the attractor space of the effective field theory. This merely pushes the threat from unconceived alternatives back a step. Whilst we can anticipate theory change *within* the attractor space, there could be an unconceived alternative outside of the attractor space of the effective field theory altogether. ? uses a historical example from ? to demonstrate this point. Newton’s law of universal gravitation (LUG) came into conflict with astronomical data in the late 19th century. Hypothetically, scientists could have formed a (highly predictively successful) modified effective theory, modifying LUG and incorporating the astronomical data. This theory would have flowed to the original LUG under a process highly analogous to renormalization group flow. However, the actual theory that replaced LUG, general relativity, lies outside of any such attractor space. It is built on different assumptions, and satisfies a different set of symmetries to LUG and the effective theory. As ?, page 1187 puts it, “The concern is that even explicit RG results are only as reassuring as the space of theories on which the RG group acts is comprehensive”.

To put it another way, the sloppy models program seems to guard us against certain kinds of doubt, but not against others. Insofar as we are not certain of the values of the sloppy parameters, (or have small uncertainty about the stiff parameters) then we can have faith in our sloppy models. However, insofar as we have doubt about the general form, f , that a latching model should take, then the sloppy models program cannot help us. The program provides us with a local defense against some limited kinds of uncertainty— but not a global defense against skepticism such as that posed by the argument from unconceived alternatives.

7 Sloppy explanation?

Now we must address the first question: can the sloppy success argument explain the success of science? Recalling section ??, we have seen, at least in certain limits, a sloppy system can be well-represented by an effective model, dependent on far fewer parameters. I have argued that it is often easier for us to create effective, latching models of the target system than a more complete model. An effective model, by definition, relies on fewer parameter combinations than a more complex model. For example, the effective model of the EGFR described in section ?? makes several simplifying assumptions. It treats C3G as directly influencing the Erk concentration, avoiding a series of intermediate steps. Likewise, effective models of thermodynamic systems work by statistically summarizing the underlying physics, removing the need to include microscopic degrees of freedom within the model. Plausibly, we need less knowledge of the target system in order to successfully create an effective model.

However, the sloppy success argument requires that many real-world systems are sloppy, i.e. they are well-represented by latching sloppy models, models that sit in the attractor basin of a simpler effective model under some particular MBAM dynamics. All

we have so far is a conditional— *if* we already know that a system is sloppy, *then* we have reason to believe our model in the face of certain kinds of uncertainty. Yet so far we have not seen any reason to *expect* that many real world systems would be sloppy.

The sloppy success argument has not yet given an explanation of the success of science. Rather, it has provided a useful reframing of the issue, in terms of system sloppiness. In effect, we have pushed the question of scientific success back a step. This has raised the question: can we explain why many real world systems are sloppy?

Ultimately, a defence of the sloppy success argument will lean on some form of inference to the best explanation. The hope is that the success of the sloppy modeling program licenses us to believe either some specific system (in the case of the local form of the argument), or many systems in general (in the case of the global form of the argument) are well-described by sloppy models. Framed in this way, the sloppy model formalism offers a toolkit for detailing the type of commitments that realists should make regarding sloppy systems.

This argument parallels the broader dialectic that I charted in section ???. Such arguments are unlikely to convince the skeptic. In particular, the skeptic would view such abductive arguments as explicitly vulnerable to the threat of an unconceived alternative explanation. Nonetheless, it is instructive to explore how this general argument form plays out in the specific arenas of effective field theory and sloppy models programs.

Quantum field theorists have provided certain heuristic arguments for why we should expect why many systems may look like quantum field theories. The most canonical of these is known as “Weinberg’s folk theorem” (see ?, chapters 1-5 for a technical exposition and ? for a philosophical reconstruction of the argument). As ?, page 8 puts it, “although you can not argue that relativity plus quantum mechanics plus cluster decomposition necessarily leads only to quantum field theory, it is very likely that any quantum theory that at sufficiently low energy and large distances looks Lorentz invariant and satisfies the cluster decomposition principle will also at sufficiently low energy look like a quantum field theory”. The argument is that any quantum field theory meeting three very general conditions considered particularly secure by scientists (quantum mechanics, Lorentz invariance, and a locality condition called cluster decomposition) is expected to approximate some quantum field theory model at low energies.²³

Note that Weinberg’s folk theorem does not fill in all the gaps here. It gives us

²³A little more precisely, Weinberg suggests that if these three conditions are satisfied:

1. There is a quantum mechanical characterisation of the S-Matrix formalism (the state of a system is a vector from a hilbert space, etc);
2. The S-Matrix obeys approximate Lorentz invariance at sufficiently low energy scales/large distances;
3. Cluster decomposition (stating that the the vacuum expectation value of a product of operators in two separated regions, A and B, can be decomposed into the expectation value for the product of the operators in A, multiplied by the expectation value for the product of the operators in B);

then any such theory will look like an effective field theory at sufficiently low energy scales in this specific sense. The Lagrangian density will be a Lorentz scalar functional of local fields, linear in the creation and annihilation operators $a(q)$ and $a^\dagger(q)$ (with coefficients that are smooth functions of momenta). This in turn means that we can perform a Fock space decomposition of the free fields.

some reason to believe that real physical systems can be approximated by quantum field theories, but even if we accept this, they need not necessarily fall into the attractor basin of our current best effective field theories. Nonetheless, the theorem provides some justification for the approach taken by physicists to pursue maximally general models within the quantum field theory framework. Even if we expect the “true” theory to not be a quantum field theory, we have some reason to expect it to approximate our quantum field theories at the energy scales we can currently experimentally probe. However, the antirealist is unlikely to be convinced. The skeptic has two obvious responses: to retreat back a level and propose skepticism about Weinberg’s three conditions, or to suggest that the “true” theory may indeed be a counter-example. Notably Weinberg’s folk theorem is explicitly vulnerable to unconceived counter-examples. The theorem does not have a proof; rather the claim is that it is extremely hard to think of a theory that defies the theorem²⁴.

Does an analogous argument hold in the case of the sloppy models program? As I explained in section ??, the eigenvalues of the FIM of sloppy models form an approximately exponential distribution. This suggests that the common features of sloppy models may indicate that they belong to a common universality class. In particular, a Vandermonde ensemble of multiparameter nonlinear models will exhibit the features of sloppy models in the limit that the system size approaches infinity²⁵. ? suggest that perhaps many real world systems can be expressed in terms of such an ensemble. However, even if this is the case, this merely sets the question back one step further. Instead of asking why many real world systems are sloppy, we ask why such real world systems can be expressed in terms of a Vandermonde ensemble.

A deeper, but closely related reason why many systems may be sloppy comes from approximation theory (?). Assuming certain smoothness conditions on the model predictions, as inputs (experimental controls such as time, experimental conditions, and so forth) are varied, then constraints can be put on the geodesic lengths of the model manifold. More precisely, if a model $y(m)$ is approximated by a Taylor series or Chebyshev series truncated to order N , for which each of the coefficients are bounded by an N -sphere of radius r , then it can be shown that the model manifold is constrained by a hyper-ellipsoid, for which the j^{th} largest principal axis length is bounded by a power law ρ^{-j} , for some $\rho \in \mathcal{N}$ (see appendix ??)²⁶. In effect, this demonstrates a connection between model smoothness and sloppiness: we should expect any model obeying some fairly general smoothness conditions to be sloppy. Even so, this pushes back the

²⁴Indeed certain counterexamples do exist. String theory is sometimes cited as one such counterexample, though its status with regards to Weinberg’s folk theorem is contentious (??).

²⁵A Vandermonde matrix is an $m \times n$ matrix in which each row may be expressed as a geometric progression, $V_{ij} = x_i^{j-1}$. ? show that, under certain assumptions, then if a model has a the Hessian matrix of the form, $H = V^T A^T A V$, where V is a Vandermonde matrix and A is a model-specific matrix, then the model will be sloppy, according to the *coordinate-dependent* definition.

²⁶?, supplemental materials suggest that a model manifold exhibiting an algebraic decay of lengths should count as sloppy. Furthermore, note, that this does not explain *all* cases of sloppiness. The argument gives a sufficient, but not necessary condition on sloppiness. The argument does not extend to probabilistic models such as the Ising model or the model for the cosmic microwave background radiation after the Big Bang, which also form hyperribbons.

question a step further: why should we expect many real-world systems to exhibit these smoothness requirements?

Compare this to Weinberg's folk theorem. Weinberg's folk theorem closes some of the gaps in the effective field theory program— using very widely accepted principles, it gives some reason to believe that many physical systems can be approximately described by quantum field theories. The sloppy models program has likewise produced a set of reasonably general conditions under which sloppiness must arise. These results go some way: insofar as these conditions are met, we should expect real world systems to be sloppy. However, it is not completely clear how often we should expect such conditions to be satisfied by the real world systems that we study. The results explicate the sloppiness in terms of some more general principles, but the explanatory gap has not been closed.

So the sloppy models program provides a partial explanation, as it currently stands, for the general success of science. It has certainly achieved a (potentially very useful) reframing of the question. Instead of asking why science is successful, given the incredible complexity of many real world systems, we can instead ask why it is that many real world systems seem to be sloppy. Or, pushing back a further step, we might ask, why it is that many real world systems are representable in terms of a Vandermonde ensemble. Pushing back further still we might ask why it is that many real world systems can be well-represented by models that obey certain smoothness constraints. This reframing might be very helpful: it gives us useful directions to pursue further research in answering the question about the success of science. However, it does not, yet, bring us all the way towards an answer.

8 Conclusions: the epistemic virtues of the sloppy models program

The sloppy models program is undeniably successful, generating many fruitful results. However, I have argued that it can, at best, give us only a partial explanation for the success of science, or for the success of particular scientific theories. The sloppy models program does help us to reframe a question about the success of science. Instead of asking, why science is successful, we might ask: why might many real-world systems be sloppy? This is a narrower, and somewhat more clearly defined question. It encourages to seek explanations for why many real world systems might have a latching model in which the set of eigenparameters follows an approximately exponential distribution. We can even push this question back a step further: why might many real world systems be representable by a Vandermonde ensemble? Even expressing a question in more clearly defined terms might be the first step towards finding an answer.

Sloppy models proponents have even offered a partial answer here. We should expect certain systems to be sloppy if they can be described by equations obeying fairly general smoothness and analyticity properties. However, it is not clear how often we should expect real world systems to be characterised by such equations, or even how one might begin to find out. In this respect, it is in a comparable position to the effective field theories program: proponents of renormalization group realism can provide some heuristic arguments that we should expect many real world systems may be approximable by

effective field theories.

How strong a support does this offer the realist? The realist might very reasonably contend that any explanation for the success of science from within science itself must stop at some point. The sloppy models program can help strengthen their case, even if it falls short of being an airtight defence against all forms of skepticism. However, the antirealist is likely to be unmoved by such claims. Whilst sloppy models provide a defence against some kinds of theory change, the crux of the argument has been about theory change to an unconceived alternative. It is precisely here that the sloppy models program seems most vulnerable.

In this sense, the debates around the sloppy models and the effective field theory programs have served as proxy conflicts in the wider war between realists and antirealists. The form of the specific arguments in each case has closely paralleled the more general arguments. The realist arguments ultimately lean upon an inference to the best explanation. These arguments are vulnerable to precisely the same antirealist challenges in the specific cases as in the general case.

Regardless, the explanation offered by the sloppy models program might at best only justify a very minimal kind of scientific realism. Once again, there is a close parallel here with the renormalization group realism program. However, I have argued that the sloppy models approach does not allow us to guard our models from all kinds of uncertainty, in particular, sloppy models might be vulnerable to unconceived alternatives. It seems likely that it is impossible to protect any scientific theory from such a challenge. However, sloppy models do exhibit a kind of robustness against certain forms of theory-change. In particular, the theories are robust against changes in the values of the sloppy parameters²⁷. Researchers in the sloppy models program envisage their theories as lying within a wider space of theories and are well aware of the limitations of their knowledge and the lessons from history. They consciously try to anticipate what types of theory change might take place, and to generate theories robust against this possibility, as far as they can. They are well aware that such theories are, in some sense, an interpolation, but one that can offer epistemic goods, both predictions and explanations. Such theories may or may not be literally true, but offer something valuable nonetheless: a kind of provisional robustness against certain kinds of theory change.

Indeed, this might also be a helpful way to view the effective field theories program. Like the case of sloppy models, I have argued that effective field theories are vulnerable to theory change to an unconceived alternative. Rather than arguing that the program provides grounds for epistemic scientific realism, it may be more convincing to argue that it provides a kind of provisional robustness against certain kinds of theory change.

This raises the question, what epistemic benefits can the sloppy models program offer? The sloppy models program provides a unifying framework, through which we can understand a wide variety of different intertheoretic reductions. We can use the MBAM approach to understand both coarse-graining, for example in a generalized Ising model, at least some cases of the renormalization group method, and other systems such as in systems biology. Better understanding how this unified family of reductions could

²⁷For a discussion of this sense of robustness, see ?

prove to be a highly fruitful direction for further research.

Appendix A The Fisher information matrix of a sloppy model

Let us suppose that we have experimentally extracted some finite number of measurements, which we represent by a set of real numbers y_m , indexed by points m . Recall that a model is a function $f : \mathcal{R}^N \rightarrow \mathcal{R}^M$ from a set of N real number parameters $\theta = \theta_n$, to a set of M real number predictions about the system, $f(\theta_n)$. Then the likelihood function, $\mathcal{L}(\theta | y_m)$ gives the joint probability distribution of observing the data y_m given the model parameters θ . Thus the likelihood function quantifies how likely it is to obtain the observed data under the given model and the given parameter values. In practice, it is often more convenient to use log likelihood function, $\log \mathcal{L}(\theta | y_m)$.

The Fisher Information Matrix gives the expectation of the second-order partial derivatives of the log-likelihood function of the observed data with respect to the model parameters. Expressed as a tensor of type (0,2), this is given by,

$$g_{\mu\nu}(\theta) = \mathbb{E}_m [\partial_\mu \log \mathcal{L}(\theta | y_m) \partial_\nu \log \mathcal{L}(\theta | y_m)] \quad (2)$$

where $\mu, \nu = 1, 2, \dots, N$, and $\mathbb{E}[\cdot]$ denotes the expectation with respect to the distribution of the m observed measurements. Often, it is more convenient to write this in terms of a suitably chosen cost function, such that the FIM is the the negative of the least squares Hessian Matrix of the cost function.

These second derivatives capture the curvature of the log-likelihood function in parameter space. High curvature indicates that the log-likelihood function is sensitive to changes in the corresponding parameter combination, whereas low curvature suggests that the log-likelihood function is less sensitive. One way to understand the FIM is as quantifying the expected information that the observed data y carries about the model parameters θ .

The eigenvectors represent the directions in which the curvature is maximal and minimal, and the corresponding eigenvalues represent the magnitudes of the curvature in these directions. In particular, the eigenvectors corresponding to the largest and smallest eigenvalues represent the principal directions of curvature, and the corresponding eigenvalues represent the principal curvatures.

Thus, geometrically, the eigenvectors of the FIM represent the principal directions in the parameter space along which the curvature of the log-likelihood function changes the most. These directions correspond to linear combinations of the original parameters. Larger eigenvalues indicate that the log-likelihood function has a high curvature along the corresponding eigenvector direction in the parameter space. This means that small changes in the parameters along this direction lead to relatively large changes in the log-likelihood, indicating higher sensitivity of the model to the changes in the corresponding parameter combination. On the other hand, smaller eigenvalues indicate that the log-likelihood function has a low curvature along the corresponding eigenvector direction in the parameter space. This means that small changes in the parameters along this

direction lead to relatively small changes in the log-likelihood, indicating lower sensitivity of the model to the changes in the corresponding parameter combination.

Interpreting $g_{\mu\nu}$ as a metric, we can define infinitesimal distances between θ and $\theta + d\theta$ over the data manifold by,

$$ds^2 = \sum_{\mu\nu=1}^N g_{\mu\nu} d\theta^\mu d\theta^\nu. \quad (3)$$

The Fisher proper length, integrated along some path $\theta(\tau)$, parametrized by τ between 0 and 1, is given by

$$L = \int_0^1 d\tau \sqrt{\sum_{\mu\nu} g_{\mu\nu}(\theta(\tau)) \frac{\partial\theta^\mu}{\partial\tau} \frac{\partial\theta^\nu}{\partial\tau}}. \quad (4)$$

Appendix B Sloppiness and approximation theory

Following ?, let us summarize the relationship between sloppiness and bounding approximations. Consider a model, $y_\theta(t)$, analytic in t , taking input parameters $\theta = (\theta^1, \dots, \theta^K)$, for some integer, K , and giving a finite-dimensional output. However, the arguments here can be extended to non-analytic functions, with higher dimension output (?, supplemental materials). We can assume $t \in (-1, 1)$ without loss of generality, if we allow the model to be appropriately transformed. Let us label the model predictions $Y_k = y_\theta(t_k)$ at N points t_k , for $N \geq K$. Now, the model manifold, \mathcal{Y} is defined as the K -dimensional surface, parametrized by $Y(\theta) = (Y_0, \dots, Y_{N-1})$ embedded in the N -dimensional prediction space.

Now, let us consider some polynomial approximations to y_θ . Let $\{\phi_j\}_{j=0}^\infty$ be a complete polynomial basis. Furthermore, let us suppose that our model has a convergent expansion in this polynomial basis, and can be written $y_\theta(T) = \sum_{j=0}^\infty b_j(\theta)\phi_j(t)$, for some coefficients $\{b_j\}$. Now, let us consider the model approximation truncated to order N , and set $\mathbf{b} = (b_0, \dots, b_{N-1})^T$. Now suppose that the parameter space is bounded by an N -sphere of radius r , i.e. $\|\mathbf{b}\| < r$. We can understand this as a smoothness condition on the model function.

Let \mathcal{P} give the manifold of the truncated approximate model. The corresponding model manifold will be distorted into a hyperellipsoid, H_P . Let us denote the linear map from parameter space to prediction space by X , $X = \phi_{j-1}(t_{i-1})$. If $l_j(H_P)$ is the diameter of the j^{th} largest cross-section of H_P then,

$$l_j(H_P) = r\sigma_j(X), \quad (5)$$

where σ_i give the singular values of X . When X has rapidly decreasing singular values, then H_P will take a hyperribbon structure. Accounting for the error in the truncated approximation, we can give hyperellipsoid bounds on the model manifold.

? considers two cases: a basis function of Chebyshev polynomials and a monomial expansion (given a Taylor series approximation of y_θ). In the Chebyshev expansion

case, we can show that the model manifold is bounded by a hyper-ellipsoid, for which the j th largest principal axis length is bounded by a power law of the form ρ^{-j} , for some $\rho \in \mathcal{N}$. In other words, the manifold of a model obeying the smoothness condition for a bounded Chebyshev expansion must have rapidly decreasing bounds on its geodesic lengths: an indication of sloppiness according to the criterion in section ???. In the monomial expansion case, we can likewise set constraints, in terms of the eigenvalues of the Vandermonde matrix.

References

- Abramowitz, G. and Gupta, H. (2008). Toward a model space and model independence metric. *Geophysical Research Letters*, 35(5).
- Apgar, J. F., Witmer, D. K., White, F. M., and Tidor, B. (2010). Sloppy models, parameter uncertainty, and the role of experimental design. *Mol. BioSyst.*, 6:1890–1900.
- Bain, J. (1998). Weinberg on qft: Demonstrative induction and underdetermination. *Synthese*, 117(1):1–30.
- Binney, J., Dowrick, N., Fisher, A., and Newman, M. (1992). *The Theory of Critical Phenomena: An Introduction to the Renormalization Group*. Oxford Science Publ. Clarendon Press.
- Brown, K. S., Hill, C. C., Calero, G. A., Myers, C. R., Lee, K. H., Sethna, J. P., and Cerione, R. A. (2004a). The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Physical Biology*, 1(3):184–195.
- Brown, K. S., Hill, C. C., Calero, G. A., Myers, C. R., Lee, K. H., Sethna, J. P., and Cerione, R. A. (2004b). The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Physical Biology*, 1(3):184.
- Brown, K. S. and Sethna, J. P. (2003). Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E*, 68:021904.
- Butterfield, J. (2014). Reduction, Emergence and Renormalization. *J. Philos.*, 111:5–49.
- Butterfield, J. and Bouatta, N. (2014). Renormalization for philosophers. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 104.
- Daniels, B. C., Chen, Y.-J., Sethna, J. P., Gutenkunst, R. N., and Myers, C. R. (2008). Sloppiness, robustness, and evolvability in systems biology. *Current Opinion in Biotechnology*, 19(4):389–395. Protein technologies / Systems biology.
- Dizadji-Bahmani, F., Frigg, R., and Hartmann, S. (2010). Who’s afraid of nagelian reduction? *Erkenntnis (1975-)*, 73(3):393–412.
- Duncan, A. (2012). *The Conceptual Framework of Quantum Field Theory*. Oxford University Press, Oxford, UK.
- Fraser, J. D. (2018). Renormalization and the formulation of scientific realism. *Philosophy of Science*, 85(5):1164–1175.
- Fraser, J. D. (2020a). The real problem with perturbative quantum field theory. *British Journal for the Philosophy of Science*, 71(2):391–413.
- Fraser, J. D. (2020b). Towards a realist view of quantum field theory. In French, S. and

- Saatsi, J., editors, *Scientific Realism and the Quantum*, pages 276–292. Oxford University Press, Oxford, UK.
- Frederiksen, S. L., Jacobsen, K. W., Brown, K. S., and Sethna, J. P. (2004). Bayesian ensemble approach to error estimation of interatomic potentials. *Phys. Rev. Lett.*, 93:165501.
- French, S. and Saatsi, J., editors (2020). *Scientific Realism and the Quantum*. Oxford University Press, Oxford, UK.
- Geris, L. and Gomez-Cabrero, D. (2016). *Uncertainty in Biology: A Computational Modeling Approach*, volume 17.
- Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., and Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLOS Computational Biology*, 3(10):1–8.
- Kitcher, P. (1993). *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford: Oxford University Press.
- Koberinski, A. and Fraser, D. (2023). Renormalization group methods and the epistemology of effective field theories. *Studies in History and Philosophy of Science*, 98:14–28.
- Laudan, L. (1981). A confutation of convergent realism. *Philosophy of Science*, 48(1):19–49.
- Machta, B. B., Chachra, R., Transtrum, M. K., and Sethna, J. P. (2013). Parameter space compression underlies emergent theories and predictive models. *Science*, 342(6158):604–607.
- Maddy, P. (2007). *Second Philosophy: A Naturalistic Method*. Oxford: OUP.
- Mannel, T. (2004). Effective Field Theories in Flavor Physics. *Springer Tracts Mod. Phys.*, 203:1–175.
- Miller, M. (2017). The structure and interpretation of quantum field theory.
- Nagel, E. (1961). *The structure of science*. London: Routledge and Keagan Paul.
- Oda, K., Matsuoka, Y., Funahashi, A., and Kitano, H. (2005). A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology*, 1:2005.0010 – 2005.0010.
- Parker, W. S. (2020). Model evaluation: An adequacy-for-purpose view. *Philosophy of Science*, 87(3):457–477.
- Polchinski, J. (1984). Renormalization and Effective Lagrangians. *Nucl. Phys. B*, 231:269–295.
- Press, W. H., S. A. Teukolsky, W. T. V., and Flannery, B. P. (1996). *Numerical Recipes in C*. Cambridge University Press, New York, 2nd edition.
- Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. Routledge.
- Putnam, H. (1979). *Mathematics, Matter and Method: Philosophical Papers*, volume 1. Cambridge University Press, 2 edition.
- Quinn, K. N., Abbott, M. C., Transtrum, M. K., Machta, B. B., and Sethna, J. P. (2021). Information geometry for multiparameter models: New perspectives on the origin of simplicity.
- Quinn, K. N., Wilber, H., Townsend, A., and Sethna, J. P. (2019). Chebyshev approxi-

- mation and the global geometry of model predictions. *Phys. Rev. Lett.*, 122:158302.
- Raju, A., Machta, B. B., and Sethna, J. P. (2018). Information loss under coarse graining: A geometric approach. *Physical Review E*, 98(5).
- Rivat, S. (2021). Effective theories and infinite idealizations: A challenge for scientific realism. *Synthese*.
- Rosaler, J. and Harlander, R. (2019). Naturalness, wilsonian renormalization, and ?fundamental parameters? in quantum field theory. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 66:118–134.
- Ruetsche, L. (2018). Renormalization group realism: The ascent of pessimism. *Philosophy of Science*, 85(5):1176–1189.
- Ruetsche, L. (2020). Perturbing realism. In French, S. and Saatsi, J., editors, *Scientific Realism and the Quantum*, page 293–314. Oxford University Press, Oxford, UK.
- Saatsi, J. (2017). Replacing recipe realism. *Synthese*, 194(9):3233–3244.
- Schaffner, K. F. (1967). Approaches to reduction. *Philosophy of Science*, 34:137–147.
- Sethna, J. P., Bierbaum, M. K., Dahmen, K. A., Goodrich, C. P., Greer, J. R., Hayden, L. X., Kent-Dobias, J. P., Lee, E. D., Liarte, D. B., Ni, X., Quinn, K. N., Raju, A., Rocklin, D. Z., Shekhawat, A., and Zapperi, S. (2017). Deformation of crystals: Connections with statistical physics. *Annual Review of Materials Research*, 47(1):217–246.
- Stanford, P. (2000). An antirealist explanation of the success of science. *Philosophy of Science*, 67:266.
- Stanford, P. K. (2003). Pyrrhic victories for scientific realism. *Journal of Philosophy*, 100 (11):553–72.
- Stanford, P. K. (2010). *Exceeding Our Grasp*. Oxford University Press USA.
- Stanford, P. K. (2021). Realism, instrumentalism, particularism: A middle path forward in the scientific realism debate. In Lyons, T. D. and Vickers, P., editors, *Contemporary Scientific Realism: The Challenge from the History of Science*. Oxford University Press.
- Transtrum, M., Machta, B., Brown, K., Daniels, B., Myers, R., and Sehtna, J. (2015). Sloppiness and emergent theories in physics, biology and beyond. *J. Chem. Phys.*
- Transtrum, M. K., Machta, B. B., and Sethna, J. P. (2010). Why are nonlinear fits to data so challenging? *Phys. Rev. Lett.*, 104:060201.
- Transtrum, M. K., Machta, B. B., and Sethna, J. P. (2011). Geometry of nonlinear least squares with applications to sloppy models and optimization. *Phys. Rev. E*, 83:036701.
- Transtrum, M. K. and Qiu, P. (2014). Model reduction by manifold boundaries. *Phys. Rev. Lett.*, 113:098701.
- Votsis, I. (2011). The prospective stance in realism. *Philosophy of Science*, 78(5):1223–1234.
- Wallace, D. (2006). In defence of naivet : The conceptual status of lagrangian quantum field theory. *Synthese*, 151(1):33–80.
- Waterfall, J. J., Casey, F. P., Gutenkunst, R. N., Brown, K. S., Myers, C. R., Brouwer,

- P. W., Elser, V., and Sethna, J. P. (2006). Sloppy-model universality class and the vandermonde matrix. *Phys. Rev. Lett.*, 97:150601.
- Weinberg, S. (1995). *The Quantum Theory of Fields*, volume 1. Cambridge University Press.
- Weinberg, S. (1996). What is quantum field theory, and what did we think it is? In *Conference on Historical Examination and Philosophical Reflections on the Foundations of Quantum Field Theory*, pages 241–251.
- Wells, J. D. (2012). *Effective Theories in Physics, From Planetary Orbits to Elementary Particle Masses*. Springer, Berlin, Heidelberg.
- White, A., Tolman, M., Thames, H. D., Withers, H. R., Mason, K. A., and Transtrum, M. K. (2016). The limitations of model-based experimental design and parameter estimation in sloppy systems. *PLOS Computational Biology*, 12(12):1–26.
- Williams, P. (2019). Scientific realism made effective. *British Journal for the Philosophy of Science*, 70(1):209–237.
- Wilson, K. G. and Kogut, J. (1974). The renormalization group and the ϵ expansion. *Physics Reports*, 12(2):75 – 199.
- Wilson, M. (2006). *Wandering Significance: An Essay on Conceptual Behaviour*. Oxford: Clarendon Press.
- Worrall, J. (1989). Structural realism: The best of both worlds? *Dialectica*, 43(1-2):99–124.
- Wu, J. (2021). Explaining universality: Infinite limit systems in the renormalization group method. *Synthese*, (5-6):14897–14930.
- Zhao, K. (2018). A statistical learning approach to a problem of induction.