Explanations and Candidate Explanations in Physics

Martin King

DRAFT DO NOT CITE OR DISTRIBUTE

Abstract

There has been a growing trend to include non-causal models in accounts of scientific explanation. A worry addressed in this paper is that without a higher threshold for explanation there are no tools for distinguishing between models that provide genuine explanations and those that provide merely potential explanations. To remedy this, a condition is introduced that extends a veridicality requirement to models that are empirically underdetermined, highly-idealised, or otherwise non-causal. This condition is applied to models of electroweak symmetry breaking beyond the Standard Model.

1 Introduction

Many scientists take themselves to be in the business of searching for and providing scientific explanations of natural phenomena. What the nature, structure, and essential features of these scientific explanations are has been a long and fruitful research tradition in the philosophy of science. It has been proposed that explanations are certain kinds of deductive arguments from law-like statements (Hempel and Oppenheim, 1948), that they capture mechanisms (Machamer et al., 2000), or latch onto the causal dependency relations in the real world (Salmon, 1984; Strevens, 2008; Woodward, 2003), and so on. Many of these approaches have featured prominent roles for some kind of veridicality requirement on explanation, generally in terms of the truth, or approximate truth, of the statements or in the representational accuracy of the model being referenced. Causal accounts of explanation are widely popular and there are many different approaches to determining the relevant causes, but typically the veridicality requirement is that the dependency relations in the model reflect the actual causes responsible for the phenomenon to be explained. A major issue for causal accounts is to properly specify which causes are the relevant ones for explaining their effects.

Recently, however, there has been a growing trend to reflect the varied and non-veridical practices of scientific modelling. Models are abstract and idealised and there are many, sometimes conflicting, epistemic aims that guide the model builder. That all models of interest in scientific explanation will accurately reflect real-world causal relations and that a model will be explanatory precisely in virtue of the causes in the system are increasingly suspect assumptions. This has gone hand-in-hand with the acceptance that non-causal models can be explanatory and a broad variety of approaches to explanation have flourished in the literature as a result. Some have focused on distinctly mathematical explanations (Lange, 2016), minimal-model explanations (Batterman and Rice, 2014), structural explanations of highly-idealised models (Bokulich, 2008), non-causal counterfactual explanations (Reutlinger, 2012; Saatsi and Pexton, 2013), and more. I believe that this trend is based on legitimate criticisms that expose critical shortcomings of causal accounts, but that proponents have been taking the wrong steps to accommodate non-causal and highly-idealised models. One of my main worries and the one that will be a focus of this paper is that in order to accommodate such models, conditions on explanation have been relaxed to the point that one cannot distinguish between models that provide genuine explanations and those that provide merely potential or *candidate* explanations.

In order to analyse this distinction, I examine some of the potential explanations of mass generation in particle physics.[1] The problem is that viable candidate explanations are not known to be incorrect and so can satisfy local conditions of empirical adequacy. Instead of what I see as a loosening of requirements for explanation, I argue that a different kind of condition needs to be imposed—one that focuses on features of the explanation outside of the local relation between the model and its target system. The idea is to find a global surrogate for local veridicality conditions. Concretely, this involves requiring that an explanatory model is part of, or can be fit to, a theory that is highly confirmed.

In the following section, I introduce this further condition, which I call the *Global Confirmation Condition*, and discuss the role of theory in explanation. In Section 3, I present the case study's explanandum and review the models of electroweak symmetry breaking that provide its potential explanations. Finally, in Section 4, I argue that existing accounts of explanation either cannot accommodate a Higgs mechanism explanation or cannot distinguish between the Standard Model (SM) Higgs, which is explanatory, and other models that provide merely potential, or candidate, explanations. I argue that what makes the SM Higgs a genuine explanation of particle masses is its place in an empirically broad and highly-confirmed scientific theory, the SM electroweak theory (EW). Other symmetry breaking models, extend the SM and propose additional particle content. Because of this they are called beyond the Standard Model (BSM). This additional particle content is as yet unconfirmed and indicates that their associated BSM EW theories are not highly confirmed and thus, the models provide merely potential explanations.

[1] This paper is primarily aimed at scientific explanations in physics. I will make no claims about explanation in biology or other disciplines. However, I believe a similar exposition to a variety of case studies in different scientific disciplines could be fruitful, but it is simply outside the scope of this paper.

2 Theory in Explanation

In most contemporary accounts of explanation, there is no role whatsoever for theory. Whether or not a model is part of a unifying, highly-confirmed theory with broad empirical scope, is simply not relevant to whether a model or generalisation is explanatory. Typically, what matters is whether the model exhibits certain kinds of features shared by the target system, so-called *local* features (Wayne, 2017). Contemporary accounts widely considered successful have focused exclusively on local conditions and have been called *common features* accounts (Batterman and Rice, 2014). These local conditions concern the relation between the model and its target system, such as the accurate representation of causal dependency relations, the ability to support a range of counterfactuals, and such other features. Throughout the paper, when I say that a model satisfies local conditions, I mean the following:

> *Local Counterfactual Condition*: an explanatory model M provides counterfactual information that shows how the explanandum E depends on M and initial, boundary, and auxiliary conditions C.

I formulate the condition in terms of counterfactuals for reasons of generality and reasons discussed in Section 4.1. These local conditions are contrasted with *global* conditions for explanation, which stem from the model's relation to a scientific theory. The problem I am addressing in this paper is that existing accounts require only local conditions, but in many cases this is not sufficient to distinguish between models that are explanatory and those that offer merely potential explanations. Let me first address why a global condition would be relevant to making this distinction.

Hempel's account of explanation is notoriously fraught with difficulty, but it featured a role for theory that has been under-emphasised since (Hempel and Oppenheim, 1948). Theory is important for explanation and for our understanding; the more we unify facts under theory, the better we understand them. As Hempel himself says, "what scientific explanation, especially theoretical explanation, aims at is... an objective kind of insight that is achieved by a systematic unification, by exhibiting the phenomena as manifestations of common, underlying structures and processes that conform to specific, testable, basic principles" (Hempel, 1966, p. 83). Friedman and Kitcher had unification play a very strong role in explanation, but it is in my esteem only one of the explanatory benefits of a model's connection a theory (Friedman, 1974; Kitcher, 1981, 1989). Kitcher's unificationism is not without its own drawbacks. It may well be intractable to provide the kind of independent, syntactic account of the explanatory power of theories that Kitcher was seeking and this is certainly not the aim here. However, this failure does not imply that we should abandon the idea that theory is a core aspect of our best scientific explanations.

A central feature of a good explanation is that it joins the derivation of a new phenomenon to that which we already understand; it simplifies, organizes, and relates phenomena and regularities, and so contributes to our understanding of that phenomenon as well as others. A new explanation presents a new or unexplained phenomenon in a similar manner to already understood phenomena, or in the context of an established body of scientific theory. If a model is capable of accounting for some target phenomenon, then connecting it with a theory can broaden and deepen our understanding of an already partially understood behaviour. Without this broader and deeper understanding that theory can provide, there is something missing for an explanation.

In the introduction, I claimed that an explanatory model must be 'part of or can be fit to' a highly-confirmed scientific theory. Let me now state this condition explicitly as follows:

> *Global Confirmation Condition*: an explanatory model M is a part of, or can be fit to, a highly-confirmed scientific theory T.

As far as I see it, this condition raises three further questions, which I address in turn:

1. What does it mean to be a *part of* a theory?
2. What does it mean to be *fit to* a theory?
3. What does it mean for a theory to be highly confirmed?

1. What it means for a model to be a part of a theory is relatively clear. It is particularly clear if one shares a syntactic view of theories in the sense of the logical positivists. For Carnap (1939) and Hempel (1965), a theory was a set of sentences in an axiomatic system of first order logic, and a model of a theory was what could be formulated and interpreted within that formal language. This view of theories is not widely regarded as accurate of scientific theories. Luckily, what it means for a model to be a part of a theory is also fairly clear on semantic view of theories, as described in (Giere, 1988; Suppes, 1960; van Fraassen, 1980) and many works since. On this view, a scientific theory is not an axiomatic framework, but a collection of scientific models. Thus, for a model to be a part of a theory, it would simple be a member of that collection of models (of course determining membership is not so trivial).

2. In the case study introduced in the following section, the model is simply a part of the theory, however, as I intend the condition to be broadly applicable, something must be said about cases where this connection is not so clear. We can identify cases where models are strongly or directly connected with theory (applying Newton's second law, solving the ideal gas law, etc.) and cases where there is no connection to theory (such as a data model or linear regression model). The most unclear cases are where a model may only be partially or thinly related with a theory. The satisfaction of the condition becomes much more involved when one considers what it means for a model 'to be fit to' a scientific theory. I have taken this language from Cartwright (1983), for whom explaining a phenomenon requires fitting a model that describes it into the basic framework of a theory (p. 152). Practice-based approaches to models and theories, such as (Cartwright, 1983; Morgan and Morrison, 1999) could be helpful in understanding this fit. On these approaches, models are to varying

degrees independent from both theory and data. Most models on this view are not simply models of data or models of theories, but something in between.

Broadly speaking, a model can be fit to a theory if it is compatible with it, or not incompatible with it. Scientific models are abstracted and idealised, and some highly idealised, the question of assessing compatibility is not trivial. Wayne (2018) has made significant progress on this question in looking at the role of theory in justifying the idealising assumptions of explanatory models. According to the global explanation requirement (GER) he introduces, a model can be explanatory if "no entity in the model that is essential to the explanation be physically impossible according to the relevant global theory" (p. 11). Even if I am not here focusing on justifying idealisations or even the model-theory connection itself, I will make use of this condition and say that a model can be fit to theory if it satisfies Wayne's GER condition.

3. Again, that the model is appropriately related to a scientific theory is not an issue in our case study, as the model is simply a part of the theory. The issue of connecting models to theories is largely bracketed in this paper in order to make progress elsewhere. Specifically, I aim to make progress in making a comparison of the key property of a scientific theory that I take to be analogous to traditional local requirement of veridicality, the degree of confirmation of the theory. In making a comparative analysis between the confirmation of theories, I can avoid having to present a quantitative degree of confirmation, or in making an analysis of the confirmation of a theory in isolation. In our case, and in many cases, it is clear that one theory is much more highly confirmed than another. In cases where this is not clear, of course a closer analysis is required to make a judgement. While difference in the Higgs case is clear enough that a short exhibition will suffice, I nonetheless present a longer and more detailed analysis of the comparative confirmation distinction of the Higgs models in XXXX [self-reference omitted]. If one were to undertake an analysis in isolation of other models, then this would involve determining how highly confirmed is highly confirmed enough. In which case, however, the absence of alternative candidate explanations may speak strongly in favour of what is available.For now, I turn to introduce the explanandum phenomenon and the models of electroweak symmetry breaking offering potential explanations.

3 Case Study: Electroweak Symmetry Breaking

Six years after the 2012 discovery of the Higgs boson, there is still a large number and variety of viable models in particle physics that offer potential explanations of particle masses via electroweak symmetry breaking. All of these models are, to some degree, still active research avenues in particle physics, as evidenced in XXXX [self-reference omitted]. The Standard Model features the simplest implementation of the Higgs mechanism, which is a single doublet with one fundamental scalar particle. This fundamental scalar particle is known as the Higgs boson and is responsible for generating the masses of the other SM particles. Other BSM models treat the fundamental scalar as part of an extended Higgs sector, or have the symmetry breaking role played by a composite, rather than a fundamental, scalar. The aim of this section is to show that although the SM is (and should be) heavily favoured as the explanation of particle masses, accommodating or singling out this explanation remains a problem for existing philosophical accounts of scientific explanation.

3.1 Target Explanandum: Particle Masses

The explanandum stems from the search for the unification of electromagnetic and weak forces in the 1950s and 60s. Because gauge theories had been so successful and had many desirable properties, physicists were looking to find a gauge theory for the photon and the weak bosons ($W^{\pm}$). Initially, this proved difficult, because the photon γ conserves parity and is long-ranged (i.e. massless), whereas the W bosons were known to violate parity and be short-ranged (i.e. massive). This means that the symmetry of the gauge must be a hidden, or what is called a *broken*, symmetry.

The first solution, proposed by Glashow (1961), was to invoke a partial symmetry; an invariance of only part of the Lagrangian. This was a move to a larger symmetry group, SU(2) × U(1) with four gauge bosons (including a new particle, Z^0). However, at this point in time there was still no good explanation of *why* the symmetry breaking masses are there. The masses are "as yet arbitrary… without a theory of the origins of these masses, any study based upon the analogy between decay-intermediaries and photons may make use only of partial-symmetries… Although we cannot say why the weak interactions violate parity conservation while electromagnetism does not, we have shown how this property can be embedded in a unified model of both interactions"(Glashow, 1961, p. 585–6). Interestingly, this reflects a philosophical distinction between knowledge *why* and knowledge *how* and acknowledges that even though it can account for the masses, an explanation is lacking. Part of what I offer in this paper is an explanation of why this explanation was lacking, i.e. what it takes to move from a potential to a genuine explanation (Section 4.2).

One of the original authors of the Higgs mechanism papers, Tom Kibble, notes that models with spin-1 boson masses added in by hand could explicitly break the symmetry. However, these "were well known to be non-renormalizable and thus unphysical" and "… because simply adding symmetry-breaking terms destroys many of the nice properties of gauge theories"(Kibble, 2015, p. 3–4). Physicists were then looking for this to be a case of a *spontaneously broken symmetry*. The aim was to find a gauge theory to unify spin-1 and lead to a massive, parity-violating $W^{\pm}$ (and neutral Z^0)[2], while maintaining a massless, parity-preserving γ.

Higgs, Englert, and Brout, and others looked to superconductivity in order to accommodate a broken symmetry. Here, there was a known method for giving mass to a force carrier without ruining the symmetry of the whole system

[2]The $W^{\pm}$ were proposed earlier by Feynman and Gell-Mann (1958), but the Z^0 was an indeliberate consequence of the larger gauge and not explicitly in need of explanation at the time.

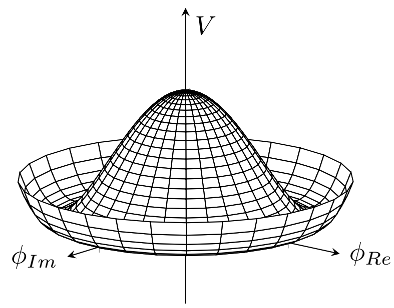

Figure 1: The Higgs potential, $V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$

(Englert and Brout, 1964; Guralnik et al., 1964; Higgs, 1964). In essence, when a gauge theory is given an additional field that breaks the symmetry, the bosons of the gauge theory acquire mass. A few years after Higgs' proposal, Weinberg (1967) applied the Higgs mechanism to Glashow's SU(2) × U(1) gauge theory. In the eletroweka symmetry breaking (EWSB) case, mass generation is induced via interaction with a symmetry breaking scalar field, known as the Higgs field. The 'Mexican hat' shape of the field's potential (as shown in Figure 1) demonstrates some of field's special properties. Because the whole potential is symmetrical, yet its lowest energy is not at the origin, the symmetry is what is called 'hidden', or 'broken'. At the lowest value of the potential V, the field has a non-zero ϕ. This what is called a vacuum expectation value (VEV), and means that the particle has an expected value everywhere in the vacuum, i.e. throughout space. When the other particles interact with this heavy field, they gain masses proportional to the strength of their interactions with the field, such that light particles interact only weakly with the Higgs, and heavy particles interact more strongly (for a standard textbook derivation, see (Peskin and Schroeder, 1995)).

The Higgs mechanism is able to account for masses within a gauge theory that provides a means to unify the electromagnetic and weak forces. Glashow, Weinberg, and Salam suspected that the theory was renormalizable, but couldn't prove it. This was only done in 1971 by 't Hooft and this step dramatically increased the viability of electroweak theory as a research programme and, I will argue, strongly reinforced claims that the Higgs mechanism explains particle masses. What is now the *SM Higgs* has come from this implementation of the simplest Higgs gauge representation. In the SM, the Higgs mechanism is realised by a single doublet that results in one physical Higgs boson—this is the model of EWSB I will refer to as the SM Higgs. The following subsection mentions some of the BSM models that feature more complex EWSB mechanisms.

3.2 Candidate Explanations

At the end of the day, it is an empirical matter as to which potential explanation, if any, is the actual reason why particles have masses. However, when it comes to letting the empirical data decide between competing explanations, we are currently at a bit of an impasse. Because of all the data from the LHC, we do know that there is a SM-like Higgs boson at 125GeV that is CP-even and neutral. We know that the particle has the non-universal couplings that one expects of SM Higgs[3], and despite record breaking amount of data-taking at the LHC, there are no significant deviations from SM expectations (ATLAS and CMS Collaborations, 2015). Many models of EWSB are being squeezed out of the remaining parameter space, some fairly strongly (e.g. see (Bechtle et al., 2016)). As is well known from arguments against falsificationism, the data will never completely rule out these possibilities. Further, any convincing or significant restrictions on some of the potential explanations at high energies may be a long time coming. Even by the end of its final run around 2035, the LHC's most sensitive searches will leave blind spots for BSM physics in the EWSB sector. We cannot conclusively say which structures of candidate explanations are actually present and which are not. This is what leads us to the recent state of particle physics with many distinct and viable models providing potential explanations of mass.

We can divide many EWSB models into different categories based on their method of breaking the symmetry.[4]

1. Higgs mechanism
   (a) SM Higgs
   (b) Extended Higgs sector
   (c) Supersymmetrically-extended Higgs sector
2. Dynamical symmetry breaking
   (a) Technicolor
   (b) Composite Higgs
3. Other (incl. extra-dimensions)

Models that feature the Higgs mechanism have the symmetry broken by the vacuum expectation value of a fundamental scalar particle, just as I described above. The SM Higgs boson is a component of a single doublet, whereas in extended Higgs sector models, the boson of EWSB is an element of larger sector, which may have two doublets, or a doublet and a singlet, and so on. Supersymmetric models require at least an additional doublet, and so are special cases of extended Higgs sectors where the particles are part of a larger symmetry group such that there are partners for all the other SM particles in addition to the extra Higgs particles. In a dynamical symmetry breaking account of masses, the analogy to superconductivity, from which Higgs and others drew insight, is

[3] The boson couples to different SM particles with different strengths, resulting in their different masses.

[4] These categories are not exclusive or exhaustive, but capture a wide range of popular models. Similar breakdowns are presented in (Borrelli and Stöltzner, 2013; Stöltzner, 2017) and XXXX [self-reference omitted].

complete. In BCS superconductivity, the symmetry is dynamically broken by a bound state of electrons, called a Cooper pair. Analogously, in the Higgs sector, the symmetry would be broken by a bound state of new heavy particles. So, instead of having a fundamental scalar to restore perturbative unitarity above the TeV scale, a new strong dynamics kicks in and instead of a light resonance, there is a tower of heavy resonances. While technicolor is widely considered dead, composite Higgs models are still pursued. Lastly, there are also EWSB models where the symmetry is broken by an appropriate choice of boundary conditions in higher dimensions.[5]

What is important in this discussion is that even though models like composite Higgs can also account for particle masses, we are not justified in claiming that a composite Higgs is the reason *why* particles have mass. Composite Higgs models may be (highly) disfavoured, but are not completely ruled out and so are at least potentially explanatory. As such, they represent exactly the kind of model we should not claim is explanatory. What all these various models offer are potential, or candidate explanations, which could be tested and promoted to genuine explanations by further research. This is not an unusual case, but typical of situations that often arise at the frontier of scientific research. For instance, several different models have been proposed to explain the Sun's anomalous coronal heating, cosmic inflation, the rotation curves of galaxies, and so on. These various models are all more or less equivalently empirically adequate and thus existing accounts of explanation would lack the tools to distinguish among them. While this paper only focuses on one case study, it is not an exceptional case, but a rather common state of affairs to have two or more viable potential explanations of some phenomenon. This situation indicates that a higher threshold for explanation is required and that is precisely what the condition I introduce provides.

Supersymmetry is often cited in physics papers as providing explanations of inflation, thermal relic abundance, the low Higgs mass, the weak hierarchy, and other outstanding issues of the SM. If supersymmetry is found to obtain, then it will be capable of explaining a lot of internal and external deficiencies of the SM—one could say that it has a great deal of explanatory promise. Though it is often claimed that such underdetermined/undiscovered models 'explain' some phenomenon, what must be meant is that they offer *potential* explanations. Otherwise, the claim that a model explains a phenomenon would be tantamount to saying that it can accommodate that phenomenon or is not incompatible with it. More than that has always been meant by 'explanation' in philosophy. Perhaps one can understand the claim as saying that the model would explain if it actually obtained. However, a model that would explain if it did obtain, *does not explain in virtue of that fact alone*, because this is true of every model. To give an explanation is to show the reason why—to show what may potentially be the reason why is only to give a potential explanation.

[5]For more on these groups of EWSB models see (Borrelli, 2012) and XXX [self-reference omitted]

4 Analysis

4.1 Applying Accounts of Explanation

It is not possible in this section to review all of the various kinds of explanations and apply them to this case with an amount of detail that would convince their proponents. However, I wish to make two claims here. The first claim is that it is far from trivial to classify the Higgs mechanism as explanatory on many accounts of explanation. In particular, it is far from obvious that the Higgs mechanism, despite its name, is a real mechanism, or that it exhibits the right kinds of causal dependency relations that would make it causally explanatory. The Higgs mechanism is not a system of entities that exhibits push-pull dynamics. But a mechanism is often considered more broadly than this. Machamer et al. (2000) have argued that a mechanism can be "entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (p. 3). If the Higgs were to fit this description, it would be an organised system that changes massless particles into massive ones. But that is simply not the case. We need a more minimal notion of what a mechanism is. Glennan (2017) has given us such a picture and he has described a 'minimal mechanism' as consisting "of entities (or parts) whose activities and interactions are organized in such a way that they produce the phenomenon" (p. 13). Even in a minimal and inclusive sense, a mechanism must have interacting parts and must involve a temporal and regular production of the phenomenon. Unlike in the Higgs-inspirational case of superconductivity, there is no underlying dynamics and no phase transition. In fact, as has been recently argued by Fraser and Koberinski (2016), the Higgs mechanism is no temporal process at all; the Higgs field is just there and so are the particle masses.

The lack of interactions and temporal order indicate that the Higgs mechanism is unlikely to be captured by causal accounts. Causal accounts of explanation gained a lot of popularity because the asymmetry that causal relations import was an obvious and effective solution to some of the major problems that plagued the D-N account of explanation (Hempel and Oppenheim, 1948; Salmon, 1989). In recent causal accounts of explanation, such as Woodward (2003), there is a shift away from causal interactions and processes. For Woodward, explanatory knowledge comes from counterfactual information about how to manipulate physical systems and learn about the causal structure of the world. He explicates his notion causation in terms of directed graphs, following (Spirtes et al., 1993) and (Pearl, 2000). One can determine whether X causes Y, by performing interventions I on X, under the certain conditions that isolate and test the effect that changing the value of the X has on the values of Y. X causes Y if the system is invariant under a range of these interventions.

In our case, how can one represent the Higgs mechanism as a directed graph? Perhaps, we could think of Y as the masses of the EW gauge bosons and X as the vacuum expectation value. This would mean that all the models have the same X. What is of more interest is the various mechanisms of the candidate explanations from which the scalar field results. X, or rather $X_1, \ldots, X_n$, should be the values of the all the parameters involved in the various EWSB models that lead to a vacuum expectation value that can generate the particle masses. Not only might this be intractable to set up as a directed graph, these are merely models and not physical systems on which one can intervene—it is not clear that there is even an I that causes X. One can relax the literal interpretation of 'manipulation' in Woodward's account, as many are inclined, but it is a stretch to include hypothetically adjusting parameters on particles that may or may not exist. Perhaps one can accommodate the Higgs mechanism into this framework, but this would require a very loose notion of causes and causal interventions. I think it is widely recognised that the prospects for a causal interpretation of the Higgs mechanism are rather slim, and that a more general framework for explanation would be better suited. Some, such as Reutlinger (2012) and Saatsi and Pexton (2013), have argued for a turn away from Woodward's causal interventions to get a general account of counterfactual explanation. My worry here is that while counterfactual accounts provide a broad and inclusive framework for explanation, there is no way to debar claims that, for instance, a composite Higgs explains why particles have masses.

This brings us to my second claim, which is that according to accounts where the Higgs mechanism can be accommodated (counterfactual accounts), there are no grounds to prefer it over merely candidate explanations. On Reutlinger's account, an explanation satisfies the following three conditions: (i) *veridicality*, (ii) *implication*, and (iii)*dependence* (Reutlinger, 2016, p. 737). (i) *Veridicality* requires that the set of generalisations $G_1, \ldots, G_m$, auxiliary assumptions $S_1, \ldots, S_n,$, and explanandum statement E are all (approximately) true. This condition sounds straightforward to assess, but all of the explanans statements about the models we are considering are at least approximately true, in that their predictions accommodate available data. It may turn out that generalisations featuring composite or heavy Higgses are not even approximately true, but at this point we just do not know. Cases like these BSM models arise specifically because the differentiating aspects of the models are underdetermined. (ii) The *implication* condition requires that the explanans (generalisations and auxiliary assumptions) logically entails the explanandum E or some conditional probability $P(E|S_1, \ldots, S_n)$, which need not be high. Given that the predictions for ranges of particle masses follows formally in the models, this condition should be trivially satisfied by the various models we are considering. (iii) Lastly, the *dependence* condition states that if S had been different, E or the conditional probability would also have been different. These models are mathematically robust physics models and the effects of changes in the values of parameters can be calculated to see what would be observed. Again, one cannot actually intervene and change the Higgs VEV or its couplings and see what actually happens. While one can only get formal results within the model, this ought to be sufficient on a non-causal account to indicate what would have happened had certain values or initial conditions been different.

The important result from applying these conditions is that other viable BSM models satisfy them just as well as the SM Higgs. The other candidate explanations are capable of accommodating a 125 GeV SM-like Higgs boson, and so they satisfy the veridicality condition as far as we can determine it. The values of the parameters of the models entail the observed values in the explanandum, to various degrees of specificity and adjusting the parameters of the models leads to differences that would be observed in the explanandum (or so the results of the model indicate). There are no tools available to this account that are capable of distinguishing between SM Higgs and various BSM models.

However, as a counterfactual account is a general framework for explanation that is broad enough to accommodate a Higgs mechanism explanation, I will here take this to form the *local condition* on explanation, which I formulated above (Section 2). Although, I believe that such a counterfactual condition would be trivially satisfied by any reasonable scientific model, this is not something I will argue here. Instead, I take the satisfaction of this condition as a basic requirement of a scientific explanation, viz. the reproduction of the explanandum and the provision of information about the model's dependency relations. I will take satisfaction of this condition as sufficient for determining a candidate or potential explanation. In the following section, I turn to the *Global Confirmation Condition* to distinguish genuine, highly-justified explanations from merely potential explanations by looking outside the derivation of explanandum from explanans, i.e. beyond the satisfaction of the local conditions.

4.2 Confirmation Comparison

The problem I have outlined so far is that existing accounts that can accommodate a Higgs mechanism explanation have no tools to distinguish between models that we ought to consider explanatory (SM Higgs) and models that merely offer potential explanations (MSSM or Composite Higgs). This is because these potentially explanatory models can satisfy any *local* conditions of veridicality (between the model and the target system) as far as they can be presently determined. My proposed solution is that a further *global* condition (between model and theory) is required. In order to see how this global condition works, we compare the confirmation of the theories that the models belong to. In the remainder of the section, I first make a compelling case that the SM Higgs is explanatory because it satisfies the global condition, and second, I argue that other symmetry breaking models are only potentially explanatory because they fail to satisfy the same condition. I will focus on the two-Higgs doublet model (2HDM), but the argument would proceed identically for any BSM model, because they all predict additional unconfirmed particle content, and it is on this that the argument hinges.

So, to proceed with the analysis, why does the SM Higgs explain particle masses? First, the model can satisfy local conditions about the explanandum, since it gives robust counterfactual information about possible changes to the system. The SM Higgs boson has precise properties that have been calculated as a function of its mass and no significant deviation from these predictions has been observed. The SM is a quantum field theory with a robust mathematical foundation and easily allows for quantitative assessments of what would have been the case, or what would have been observed, had things been different, such as had the particle had a different mass, a different spin, or coupled with different strengths to various other particles.

Importantly, it also satisfies the confirmation condition as the SM Higgs is also a part of a highly-confirmed scientific theory, in this case the SM electroweak theory. SM EW theory is so strongly confirmed at present that the particle is assumed to be *exactly* SM-like and incorporated into the background in the search for new physics (Ellis, 2017). My claim is that it is the embedding of the Higgs into a theory, and that theory's subsequent successes that distinguish it from the other candidate explanations. One can see that this is the case because Higgs' proposal was not explanatory until the EW theory it is a part of was highly-confirmed.

At the time of Higgs' proposal in 1964, there was no explanation of particle masses. There was only a candidate explanation. Even shortly later, after Weinberg incorporated the Higgs mechanism into a gauge theory of leptons, significant doubts were expressed: "Of course our model has too many arbitrary features for these predictions to be taken very seriously" (Weinberg, 1967, p. 1265). However, as the EW theory became increasingly confirmed, the more justified we became in thinking that the SM Higgs mechanism explained particle masses.

There are several notable milestones that increased the degree to which the EW theory was confirmed.The first major step was when Weinberg's theory was shown to be renormalizable by 't Hooft (1971). After that, neutral currents were predicted then observed in 1973, and the $W^{\pm}$ and Z^{0} bosons were discovered in 1983 with the masses predicted using the SM Higgs. During this time, other candidate explanations were proposed. These alternatives accounted for known observables, and also predicted SM-deviations. These deviations were not observed in the precision measurements at LEP in the 90s. Thus, even before the Higgs boson discovery, the SM Higgs was among only a few viable alternatives and as the minimal, simplest, and most austere, was (and remains) the most highly-confirmed implementation of the symmetry breaking mechanism. The main alternatives, composite Higgs and supersymmetry, were strongly favoured by many physicists, but not for the degree to which they had already been confirmed. Their popularity stemmed from their explanatory promise in making up for some of the SM's shortcomings, like offering dark matter candidates and having a naturally low Higgs boson mass.

By far the biggest jump in the justification of explanatory claims about the SM Higgs came with the 2012 discovery and has only been increasing since. As more and more data is analysed, the evidence fails to indicate any significant deviations from the SM Higgs predictions, and further, the energies being probed by the LHC are making the additional new physics of alternative proposals less likely to be found in the future. Since 2012, with the discovery of the Higgs boson, the SM EW theory has become *very* highly confirmed and we are strongly justified in claiming that SM Higgs explains particle masses.

Now consider by contrast, an other model of electroweak symmetry breaking, the two-Higgs doublet (2HDM) and its associated theory. I consider this model as it is the most minimal extension of the Higgs sector. At first glance, it looks as though this theory should also be the SM, just like in the case above. After all, the SM is modular and its 'sectors' are largely independent. One can substitute in a more complicated Higgs sector and leave the rest of the SM the same. However, a different EWSB sector indicates a different theory. A two-Higgs doublet together with the rest of the SM, would not be the SM, but some SM'. These are the theories we should compare: the SM EW theory and the 2HDM EW theory.[6]

An extended 2HDM model may be a part of a theory of some EW theory, but it is not well confirmed. The 2HDM predicts five physical Higgs boson, but no hints of the 4 additional bosons have been observed at the LHC. As mentioned above, versions of the 2HDM with minimal or constrained minimal supersymmetry have been strongly ruled out. The only aspects of these models that are well-confirmed are the SM aspects. In XXXX [self-reference omitted], I have provided a fuller and more in-depth analysis of the confirmation of various BSM models. What I argue is that because BSM models have extended particle content, none of which has been confirmed by data, the SM remains the only highly-confirmed electroweak theory. My claim here is that this provides empirical justification for claiming that it is explanatory, even though the models themselves are locally underdetermined, i.e. they satisfy local dependency conditions.

This argument is even clearer for other models of EWSB where there are many more particles, processes, and larger deviations predicted, which are more strongly disconfirmed by data. Even in so-called minimal supersymmetry, there are 105 additional parameters and more new particles than there are already in the SM, none of which have been observed. All BSM models of EWSB predict additional unconfirmed particle content in the EW sector and, as a result, these BSM EW theories are not highly confirmed.

5 Conclusion

I have introduced a global confirmation condition on explanation that ensures that an explanatory model is part of, or can be fit to, a highly-confirmed scientific theory. This condition provides a way of maintaining a kind of veridicality requirement for explanation that works when competing models are underdetermined. It does not, however, only work in this context, but would provide a means of maintaining veridicality for models that cannot be properly assessed by their representative accuracy, such as highly-idealised models and non-causal models in general. Because the emphasis is not on the common features that a model shares with its target system it is well-suited to capture a variety of idealised and highly-idealised explanatory models.

The condition I have presented is somewhat conservative in that only well-established models can be genuinely explanatory, but I have also articulated the

[6]While I think it is more accurate to compare the EW theories, the same conclusion should follow comparing the SM and some SM'.

notion of a *candidate explanation* and recognise the importance of pursuing and developing these. The SM and BSM models play different roles in explanation in particle physics. BSM models are attempts to explain the as-yet-unexplained in particle physics—they have been developed in the pursuit of further explanations. And so, as Stöltzner (2017) shows, the SM is both explanans and explanandum. A physicist can use the Higgs potential in an explanation of masses, but also pursue some BSM models that offer candidate explanations of why the Higgs potential has its characteristic shape. Explanations in particle physics have certainly not bottomed out at the SM—as anyone with children knows: there are always more why-questions. This is not challenged by my claim that the SM Higgs provides a genuine explanation but a BSM model does not—at least, not yet.

In this paper, I have argued that looking outside of the model-theory relation—beyond the local conditions—can provide a means to implement a veridicality analogue for underdetermined (and highly-idealised) models. I have presented a condition that maintains a high threshold for explanation, whose aim is not to debar pseudo-scientific models or those which are known to be inaccurate or false, but to distinguish competing, viable explanations. This is not something that is widely attempted, and I think, to the detriment of the literature. A account of explanation with a high-threshold condition can provide an important valuation of potentially explanatory models—something that I take to be a worthy goal of the philosophy of science.

Acknowledgements

[Omitted]

References

ATLAS and CMS Collaborations (2015). Combined measurement of the Higgs boson mass in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS experiments. *Physical Review Letters*, 114(19):191803. arXiv:1503.07589 [hep-ex].

Batterman, R. W. and Rice, C. C. (2014). Minimal model explanations. *Philosophy of Science*, 81(3):349–376.

Bechtle, P. et al. (2016). Killing the cMSSM softly. *Eur. Phys. J.*, C76(2):96.

Bokulich, A. (2008). *Reexamining the Quantum-Classical Relation: Beyond Reductionsm and Pluralism.* Cambridge University Press, New York.

Borrelli, A. (2012). The case of the composite higgs: The model as a "rosetta stone" in contemporary high-energy physics. *Studies in History and Philosophy of Modern Physics*, 43(3):195–214.

Borrelli, A. and Stöltzner, M. (2013). Model landscapes in the higgs sector. In Karakostas, V. and Dieks, D., editors, *Epsa11 Perspectives and Foundational Problems in Philosophy of Science*, pages 241–252. Springer.

Carnap, R. (1939). Foundations of logic and mathematics. *Journal of Philosophy*, 36(23):636–637.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford University Press, Oxford.

Ellis, J. (2017). Topics in Higgs Physics. In *3rd Asia-Europe-Pacific School of High-Energy Physics (AEPSHEP 2016) Beijing, China, October 12-25, 2016*.

Englert, F. and Brout, R. (1964). Broken symmetry and the mass of gauge vector mesons. *Physical Review Letters*, 13(9):321–323.

Feynman, R. P. and Gell-Mann, M. (1958). Theory of the fermi interaction. *Physical Review*, 109(1):193.

Fraser, D. and Koberinski, A. (2016). The higgs mechanism and superconductivity: A case study of formal analogies. *Studies in History and Philosophy of Modern Physics*, 55.

Friedman, M. (1974). Explanation and scientific understanding. *Journal of philosophy*, 71(5).

Giere, R. N. (1988). Explaining science: A cognitive approach. *Philosophical Review*, 100(4):653–656.

Glashow, S. (1961). Partial symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588.

Glennan, S. (2017). *The New Mechanical Philosophy*. Oxford UP, Oxford.

Guralnik, G., Hagen, C. R., and Kibble, T. W. (1964). Global conservation laws and massless particles. *Physical Review Letters*, 13(20):585–587.

Hempel, C. G. (1965). *Aspects of Scientific Explanation and Other Essays*. The Free Press, New York.

Hempel, C. G. (1966). *Philosophy of natural science*. Prentice-Hall foundations of philosophy series. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Hempel, C. G. and Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15(2):40.

Higgs, P. (1964). Broken symmetries and the masses of gauge bosons. *Physical Review Letters*, 13(16):508.

Kibble, T. W. B. (2015). History of electroweak symmetry breaking. *J. Phys. Conf. Ser.*, 626(1):012001.

Kitcher, P. (1981). Explanatory unification. *Philosophy of Science*, 48(4):507–531.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In Kitcher, P. and Salmon, W., editors, *Scientific Explanation*. University of Minnesota Press, Minneapolis.

Lange, M. (2016). *Because Without Cause: Non-Causal Explanations in Science and Mathematics*. Oxford University Press USA.

Machamer, P., Darden, L., and Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1).

Morgan, M. and Morrison, M., editors (1999). *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge UP.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge.

Peskin, M. and Schroeder, D. (1995). *An introduction to quantum field theory*. Harper Collins.

Reutlinger, A. (2012). Getting rid of interventions. *Studies in History and Philosophy of Biol & Biomed Sci*, 43(4):787–795.

Reutlinger, A. (2016). Is there a monist theory of causal and noncausal explanations? the counterfactual theory of scientific explanation. *Philosophy of Science*, 83(4):733–745.

Saatsi, J. and Pexton, M. (2013). Reassessing woodward's account of explanation: regularities, counterfactuals, and non-causal explanations. *Philosophy of Science*, 80(5):613–624.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press, Princeton, NJ.

Salmon, W. C. (1989). *Four decades of scientific explanation*. University of Minnesota Press, Minneapolis.

Spirtes, P., Glymour, C. N., and Scheines, R. (1993). *Causation, prediction, and search*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass., 2nd edition.

Stöltzner, M. (2017). The variety of explanations in the higgs sector. *Synthese*, 194(2):433–460.

Strevens, M. (2008). *Depth: An Account of Scientific Explanation*. Harvard University Press, Harvard, MA.

Suppes, P. (1960). A comparison of the meaning and uses of models in mathematics and the empirical sciences. *Synthese*, 12(2/3):287–301.

't Hooft, G. (1971). Renormalizable lagrangians for massive yang-mills fields. *Nuclear Physics, Section B*, 35:167–188.

van Fraassen, B. C. (1980). *The Scientific Image*. Clarendon library of logic and philosophy. Clarendon Press, New York.

Wayne, A. (2017). Explanatory integration. *European Journal of Philosophy of Science*, pages 1–19.

Wayne, A. (2018). Point-particle explanations: The case of gravitational waves. *Synthese*, pages 1–21.

Weinberg, S. (1967). A model of leptons. *Physical Review Letters*, 19:1264 – 1266.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, Oxford.