Synthese (2017) 194:461–476
DOI 10.1007/s11229-015-0941-8

S.I.: EVIDENCE FOR THE HIGGS PARTICLE

The Higgs discovery as a diagnostic causal inference

Adrian Wüthrich[1,2]

Received: 1 April 2014 / Accepted: 30 September 2015 / Published online: 13 October 2015

Abstract I reconstruct the discovery of the Higgs boson by the ATLAS collaboration at CERN as the application of a series of inferences from effects to causes. I show to what extent such *diagnostic causal inferences* can be based on well established knowledge gained in previous experiments. To this extent, causal reasoning can be used to infer the existence of entities, rather than just causal relationships between them. The resulting account relies on the principle of causality, attributes only a heuristic role to the theory's predictions, and shows how, and to what extent, data selection can be used to exclude alternative causes, even "unconceived" ones.

Keywords Principle of causality · Causal inference · Inference to the best explanation · Unobservable entities · Higgs boson · Discovery · Justification · Heuristics · Theory · Experiment · Data selection · Unconceived alternatives

1 Causal inference to "unobservable" entities

A great deal of scientific discourse is about entities which are far from being accessible by our unaided senses. They are often called "unobservable", "theoretical", or "hid-

✉ Adrian Wüthrich
adrian.wuethrich@tu-berlin.de
http://www.tu-berlin.de/?id=133444

1 Institut für Philosophie, Literatur-, Wissenschafts- und Technikgeschichte, Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany

2 Present Address: Max Planck Institute for the History of Science, Boltzmannstr. 22, 14195 Berlin, Germany

den" entities.[1] For instance, scientists claim there to be genes, electrons, gravitational fields and so on, and they even claim to have a considerable amount of knowledge about these entities. What methodology, if any, do scientists follow, implicitly or explicitly, to arrive at such claims? How can the philosophy of science account for the scientists' apparently quite successful practice? One possibility is to explain the scientists' procedure by appeal to a hypothetico-deductive method (Hempel 1966). On such an account, they assume the existence of the unobservable entity for the sake of the argument and if the implications of this assumption match pertinent empirical findings they think to have good reasons to adopt the hypothesis into their set of fairly well established scientific beliefs. A variant of this account is to say that the scientists proceed by inference to the best explanation. According to this view, they first try to come up with a pool of potential explanations for their empirical findings and then choose the best, or (to speak with Lipton 2004) "loveliest" of them. Inference to the best explanation states that the best explanation is also the most likely to be true or, at any rate, the one that the scientists have the best reasons to believe in.

A shortcoming of both the hypothetico-deductive and the best explanation account is that they heavily rely on the scientists' ability to come up with reasonable hypotheses or potential explanations in the first place. The eventually inferred hypothesis or explanation can only be as close to the truth, or as good to whatever standards you have, as the hypotheses or explanations in the initial pool. In other words, the scientists might be bound to choose the "best of a bad lot" (van Fraassen 1989, p. 143).

Also, the need for an initial pool of hypotheses or explanations makes the conclusion of this type of inferences crucially depend on the scientists' current knowledge and on their imaginative and creative capabilities. This, in turn, also makes the conclusion of such inferences liable to significant revisions since, as even a cursory look at the history of science may teach us, the scientists' current knowledge indeed often was significantly revised, and I do not see any reason why this should be different just now.

To avoid these shortcomings, we may want to look for alternative methods of inferring unobservable entities—a method, that is, which does not need a fairly articulated pool of initial hypotheses or explanations. The method should then also provide more robust conclusions in the sense that they would be less liable to significant revision. Since several authors have pointed out that causal warrant is often one of the most robust warrants for our scientific claims (Egg 2012; Suárez 2008; Cartwright 1983), causal inference methods may seem, at first sight, to be good candidates for what we are looking for.

On closer inspection though, it is not so clear whether causal methods can indeed fit the bill. Causal inference methods, be they based on a deterministic or probabilistic account of causation (Baumgartner 2009; Spirtes et al. 1993), start from data about the presence or absence of certain factors involving certain entities and infer from there which factors are causally relevant for which. For instance, you apply algorithms to determine from a set of statistical or probabilistic dependencies a set of causal dependencies—a causal structure. Or you manipulate certain factors and see how the system under investigation reacts (Woodward 2003). Clearly, all this presupposes,

[1] Cf. Arabatzis (2012, Sect. 9.2) for the difficulty of choosing an adequate term.

rather than infers, the existence of the involved entities; otherwise we could not have observed occurrences of the factors, we could not know to have manipulated them, or we could not have gathered statistical data about them.

A special case of the difficulty I try to spell out here is Lipton's problem of "inferred differences". He formulates it thus:

> The Method of Difference sanctions the inference that the only difference between the antecedents of a case where the effect occurs and one where it does not marks a cause of the effect. Here the contrastive evidence is not for the *existence* of the prior difference, but only for its causal role. The method says nothing about the discovery of differences, only about the inference from sole difference to cause. (Lipton 2004, p. 127, emph. in the original)

By the method of difference, Lipton says, you can infer for which factors an entity acts as a cause but only once it has been observed to mark a difference between sufficiently homogeneous situations. However, or so I understand Lipton, the observation of the entity already implies the existence of it. The existence of an entity, therefore, enters into the premises of the method of difference rather than into its conclusion. Also, because of the requirement that a difference be observed, the method does not seem to be applicable to unobserved, or even "unobservable", entities.

In Lipton's view then, cases where differences have to be inferred, rather than being observed, are beyond the scope of the method of difference. It is in these cases, Lipton claims, that an inference to the best explanation is indispensable. Accordingly, in such cases, the reliance on a potentially "bad" and contingent initial pool of hypotheses or explanations seems unavoidable. Lipton thus raises a challenge to show how the method of difference, or similar causal methods, can be employed to infer differences. More broadly construed, the challenge amounts to provide a method for inferring existence claims without the need of already hypothetically articulating what eventually will be inferred.

Steven Rappaport in 1996, and more recently Raphael Scholl, have taken up Lipton's challenge (Rappaport 1996; Scholl 2015). Rappaport, on the one hand, argues that John Stuart Mill's method of *residues* (rather than of differences) is a way of causally inferring unobserved differences. Scholl, on the other hand, notices that, in the important cases, the differences, although not observable in a strict sense, can be detected such that there is no need to infer them on the ground of their providing a best explanation.

In both Rappaport's and Scholl's proposals, the unobserved entities or differences are inferred, or detected, in a "diagnostic" way, that is to say, by an inference from the instantiation of a certain type of effect to an instantiation of one of its types of causes.

However, in their examples, the entity was either observable and eventually observed (like Neptune in the case Rappaport uses), or the entity was, even though unobservable, already known to exist (like the cadaveric matter in the case of Ignaz Semmelweis' investigations into the causes of childbed fever, which Scholl studies). Therefore, Rappaport and Scholl did or need not make clear how the diagnostic inference is supposed to work in a case where the mere existence of the entity has yet to be established (unlike with cadaveric matter) and where no visual observation is possible in any direct sense (unlike with the planet Neptune). Thus, for the case of a thoroughly unobservable entity, Lipton's challenge does not seem to have been met.

Just such a case is provided by the discovery of the Higgs boson in 2012 by the ATLAS and CMS collaboration at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland. I take it that elementary particles, such as the Higgs boson, are paradigm cases of what is usually meant by "unobservable" entities — even for someone who is skeptical about whether such a term is meaningful at all. Also, the Higgs boson is a paradigm case for an entity of which the existence has been, though theoretically expected, a long-standing open question. To establish its existence was one of the aims of the experiments, such as ATLAS, at the Large Hadron Collider (LHC) at CERN. With two publications by the ATLAS and CMS collaborations, following an announcement July 4, 2012, the aim has been achieved to a large extent (ATLAS Collaboration 2012; CMS Collaboration 2012).

In the remainder of the article, I will reconstruct the discovery of the Higgs boson as a case of a causal inference (Sect. 3). I will argue that the example shows how scientists are able to infer the existence of an unobservable entity without relying, hypothetically, on the prior assumption of its existence. I will end with a discussion of the possible consequences of such an account (Sect. 4).

As the title of my article indicates, I will not depart completely from Rappaport's and Scholl's more or less explicit proposals to see a diagnostic inference at work. Rather, my aim is to show how a suitable causal inference, similar to the ones they propose, can be made without already presupposing the existence of the entity in question and without the possibility of an eventual visual observation.

2 Diagnostic causal inferences

Before proceeding to the study of the case of the Higgs discovery, I will state the problem I intend to address and give an outline of my proposed solution to it in a framework of causal reasoning. To a large extent I presuppose a regularity theoretic account of causation such as the one described in Graßhoff and May (2001), Baumgartner and Graßhoff (2004), and Baumgartner (2008). Such an account is appropriate and promising in particular because of its "lightweight analytical toolbox" (Baumgartner 2013, p. 85), its match with "pre-theoretic intuitions" (Baumgartner 2008, p. 328), and its track-record of application in illuminating historical case studies (e.g., Nickelsen and Graßhoff 2011; Graßhoff and May 2003). In particular, my account of the Higgs discovery heavily relies on the *principle of causality*. Other than that, however, my analysis and interpretation does not depend significantly on the particular causal approach taken.

The principle of causality states that no type of effect is instantiated when no one type of its causes is instantiated as well. The slogan is "no effect without a cause". Note that this is logically independent of the principle of determinism, in particular, the principle of causality does not imply the principle of determinism. Even if the same cause is not always followed by the same effect, it may be the case that no effect occurs without any of its causes. Indeed, I will not need to rely on a principle of determinism, and leave open the question whether deterministic quantum theories, such as Bohm's

(1952a, b), can be empirically adequate for elementary particle processes. Also I cannot show that, and how, the basic attractive features of a regularity theoretic account of causation can be saved if the principle of determinism failed. Suffice it to say that one will probably have to formulate the regularities on the level of statistical averages rather than individual instances of event types. In fact, all this does not have to be worked out for the present purposes, because I will only rely on the assumption that any type of effect is only instantiated when one of its types of causes is instantiated as well (i.e. the principle of causality). And this can hold true even if the principle of determinism fails.

On the basis of the principle of causality, if you observe or otherwise establish the instantiation of a type of effect for which no known type of cause is instantiated, you can infer the existence of a hitherto unknown type of cause for this type of effect. Otherwise, an effect would have been instantiated without an instantiation of one of its types of causes, and the principle of causality would be violated.

I submit that such an inference constitutes by itself already an essential part of the Higgs discovery, which I will discuss below (Sect. 3). Through a sophisticated procedure of data selection, the ATLAS researchers are able to find a set of events that is very unlikely to be caused by the known particles. In other words, the probability of instantiation of that set of events given the "null hypothesis", which says that only the known causes are operating, is very low.[2]

Ideally, the researchers would find a set of events for which it is impossible, rather than only improbable, to be caused by the known causes. But due to practical limitations and quantum mechanical indeterminism, there is always a non-vanishing probability that the selected events are caused by the known particles after all. However, consensus nowadays has it that the discovery claim is justified if that probability is no greater than of order 10^{-7}. This corresponds to a statistical deviation from the mean value of the expected number of events brought about by the known processes of more than five standard deviations, or "five sigma".

Moreover, the principle of causality licenses *diagnostic causal inferences*. Suppose you know the causal relevance of a certain type of event (the cause) for another type of event (the effect). Suppose further that all other types of causes for this effect are not instantiated in the situation under consideration. You can then infer the instantiation of the particular type of cause from an instantiation of the type of effect.

For my present purposes, successful diagnostic inferences have to solve two main difficulties—one rather theoretical, the other rather practical. Let us start with the practical difficulty. While the inference from the exclusion of the instantiation of all but one type of cause, together with the principle of causality, to the instantiation of the remaining type of cause is valid, it may seem, in practice, almost impossible to exclude the alternative causes. In particular, a successful diagnostic inference does

[2] Here and in the remainder of the article, I use statements about the probability of instantiations of event types given the instantiation of certain other event types, and statements about the probability of the former instantiations being caused by the latter as equivalent. I take it that the way regularity theories of causation deal with causal relations between token events (rather than between event types) justifies this equivalence (cf. Baumgartner 2013).

not only need the exclusion of *known* alternative causes but of unknown, and even "unconceived" ones as well (cf. Stanford 2006).

While the practical difficulty is a rather general one, the theoretical difficulty is more directly related to the main problem I intend to address in the present article. The theoretical difficulty is that, when it comes to infer the existence of unobservable entities, it seems unreasonable to demand that we already know what effects the unknown entity can have. After all, to establish a causal relevance relation means to apply the method of difference, probabilistic algorithms, or interventional methods or the like. But, as mentioned in the Introduction (Sect. 1), these methods presuppose the existence of the entities involved.

It turns out, I believe, that the two problems, the practical and the theoretical one, have basically the same solution. The way out of the apparent impasses is to distinguish causal relevance relations of different specificity. The same two entities can instantiate more or less general causal relevance relations. For example, Neptune is causally relevant for Uranus' trajectory insofar as it exerts a gravitational influence on it. But the two entities also instantiate the more general relation that one accretion of matter is causally relevant for another accretion of matter as described by the law of gravitation (cf. Scholl 2015, p. 105).

If we believe that it is possible to establish such general regularities by the method of difference or variants thereof (and I think we do) we must consider it sufficient to have applied the method to only a finite (and usually small) number of instances. We cannot observe the consequences of every possible variation of the mass parameter characterizing the accretions of matter, simply because there are infinitely many variations. A finite number of such difference tests must suffice, under sufficiently homogeneous circumstances, to give good reasons to believe in the general causal relevance relation.

In this way we can have at our disposal a causal relevance relation that can be used for a diagnostic causal inference even without presupposing the existence of the entity in question. An entity can, if it turns out to exist, instantiate causal relevance relations even if the method of difference to establish the causal relevance relation was not applied to the entity itself. Instead, using well-known entities, you can establish causal relevance relations that are sufficiently general to be instantiated by the hitherto unknown entity if it turns out to exist. Thus we can use, for a diagnostic causal inference, an established causal relevance relation even for an entity of which we do not even know whether it exists. This solves the theoretical problem considered above.

But the move to more general causal relevance relations also solves the practical problem of excluding alternative causes to a significant extent. There usually are causal relevance relations which are both so general as to not leave any reasonable alternative left and, at the same time, specific enough as to allow for interesting and non-trivial diagnostic causal inferences. In the Higgs case (see Sect. 3), the conservation of relativistic energy and momentum will play the role of such a relation. If the observed effects in the particle detector are classified according to their energy and momentum, virtually the only type of cause, worthwhile considering, consists of events with values of energy and momentum such that, in the transition from cause to effect, these quantities are conserved.

Of course, it is not impossible that even laws and regularities which, at some moment, seemed completely secure turn out to be violated in some cases. The failure of the law of inverse square dependence of gravitational attraction on distance, in the attempt to account for the precession of Mercury's perihelion, is probably a case in point. It might have seemed completely unreasonable, for some time, to consider alternative causes which implied the failure of this law. Yet the cause of the precession, the general theory of relativity tells us, is the delay in the effect of the gravitational influence of the Sun on Mercury, which leads to a slight violation of the inverse square law of gravitational attraction. But I take such cases to be so rare that it is still fair to say that the conclusions presupposing very "low level" regularities, such as energy conservation, are as robust as we can possibly get.

So, by using such a general causal relevance relation that there remains virtually only one worthwhile type of cause, the problem of the exclusion of alternative causes is circumvented. Of course, the price to pay for this is that the conclusions of the resulting inferences are more general than what one may have hoped for. I will take up this issue below (Sects. 3 and 4) where I am going to have a closer look at the case of the Higgs discovery.

3 Observation of a new particle at the LHC

For the reconstruction of the ATLAS collaboration's discovery, I take as my basis the published article from August 2012 (ATLAS Collaboration 2012), which speaks of the "observation of a new particle" in its title. I will restrict myself to the case of the search for the Higgs boson in reactions leading to a final state consisting of four leptons ($4l$) through an intermediate state of two Z bosons. I do not pretend to give anything like a comprehensive reconstruction of the episode. Rather, I will argue that the description of the discovery as a causal inference is at least a plausible candidate for an adequate description of such cases. The detailed evaluation of the adequacy of such a description and a thorough comparison to other accounts would need the details of the case study, including unpublished material. All this is beyond the scope of the present article.[3]

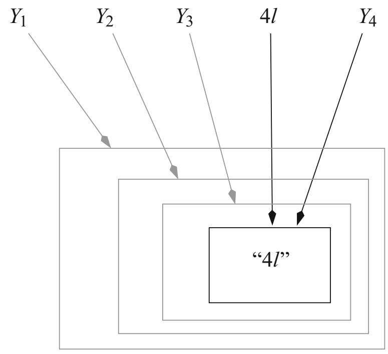
The ATLAS collaboration divides the search for the Higgs particle into several "channels". Each channel deals with the final states of one of the possible types of decay of the Higgs particle as predicted by the current theory, the Standard Model. As mentioned, I will study the case of the analysis concerned with the predicted decay of the Higgs boson through two Z bosons into four leptons, where a lepton (l) is either an electron, a positron, a muon, or an anti-muon.[4]

The first step in this analysis is the establishment of the presence of 4 leptons in the final state of one of the collisions. Their presence is inferred from their characteristic reaction with the different pieces of material in the detector. For instance, the presence

[3] For a study of the case of the discovery and detection of the W boson, which includes some unpublished material and defends a similar thesis as the one proposed here, see Wüthrich (2012).

[4] Often, "electron" (e) and "muon" (μ) is used to refer to either the particle or the anti-particle, i.e. positron or anti-muon. They have the same "flavour" but opposite charge (see, for instance, ATLAS Collaboration 2012, p. 3).

Fig. 1 Exclusion of alternative causes through data selection



of an electron is inferred from the fact that a track in the inner detector points to a deposit of energy in the electromagnetic calorimeter. The shape of the deposit must satisfy several further criteria.[5] Similar procedures are followed for the muon and the anti-muon.

This is a first instance of a diagnostic causal inference (see Sect. 2). From previous experiments, test runs and material science the researchers know how electrons and muons behave in various types of materials. The researchers can use these established causal relevance relations to infer instances of certain types of causes (the presence of electrons and muons) from instances of certain types of effects (characteristic reactions in the detector material).

Through selection procedures the researchers try to find events such that it is almost impossible that particles other than leptons would have produced the selected reactions in the detector. Data selection thus serves to exclude, as much as possible, alternative causes for certain effects. While an initial set of events (corresponding to the largest box in Fig. 1) may have been caused by instantiations of many different types of events (Y_1, Y_2, Y_3, Y_4, and $4l$, see Fig. 1), a subset (corresponding to the smallest box in the figure) of the initial set of events will usually have much fewer plausible causes (Y_4 and $4l$).

However, there are practical and theoretical limits to the exclusion of alternative causes through data selection such that, even for the final selection of events, the instantiation of some alternative types of causes (Y_4) cannot be excluded completely.

In practice the detector can only resolve so much of a difference in energy deposit or other characteristic quantities. Therefore, there will usually be a certain fraction of the selected reactions that will not be caused by leptons but instead by photons, hadronic jets, or other particles that behave similarly to leptons. Also, it is practically impossible to suppress the influence of unspecified or unknown "disturbing" factors such as faulty material, mechanical stresses, uncontrollable discharges in the apparatus etc. In this sense, photons and hadronic jets (or other events) may "look" like leptons, and they are erroneously identified as such in the analysis.

5 "Electron [or positron] candidates must have a well-reconstructed [inner detector] track pointing to an electromagnetic calorimeter cluster and the cluster should satisfy a set of identification criteria [...]" (ATLAS Collaboration 2012, p. 3).

The theoretical limits to the exclusion of alternative causes are due to the fact that the quantum mechanical reactions of the particles with the detector material are often not deterministic. The number of events caused by a certain type of cause may fluctuate, even under ideal circumstances, relative to the cause's average behavior. Therefore, the type of effect that is normally caused only by a specific type of cause (the leptons) may be caused by other causes (involving other particles) in some instances. In these instances, the diagnostic inference from the effect to the normal cause again misidentifies other particles as leptons.

At any rate, the ratio as well as its fluctuation of such misidentified events can be estimated, at least in principle, from the results of previous experiments, test runs and knowledge of material science. The reaction of the detector material to the presence of the leptons does not depend on the particular circumstance that they might be produced by a Higgs boson. The researchers can use well-known reactions, such as the production and decay of a Z boson into leptons, to investigate the way in which the leptons and the detector material interact and how often the detector misidentifies photons and hadronic jets as leptons. The well-known reactions produce a sufficiently broad spectrum of leptons such as to calibrate their reaction with the detector material even for the energies at which they will, most of the time, occur if they are to be produced by the decay of a Higgs boson. For instance, on average, the final state leptons from a Higgs decay will have a higher energy than those coming from well-known reactions. But even in the well-known reactions (at lower energies) do the kinematics allow final state leptons with energies as high as those which often occur in the hypothetical Higgs decays.[6]

The situation is illustrated by the lowest two levels of Fig. 2. The selected reactions, "$4l$", in the material of the detector, the "traces" of the leptons as it were, are most often indeed caused by the passage of four leptons, $4l$. But a certain fraction of this type of reactions is caused by photons, γ, hadronic jets, or disturbing factors. Misidentified events are one source of so-called "reducible background"; I will discuss other sources below.

Once the presence of electrons and muons is established (apart from a certain amount of misidentification), the central part of the Higgs search in this channel begins. The Standard Model predicts the decay of the Higgs boson into four leptons, $4l$, through an intermediate state of two Z bosons, one of them possibly virtual, $Z^{(*)}$.[7] To find reactions with four leptons in the final state which resulted from such an intermediate state, $ZZ^{(*)}$, several selection criteria are applied, among them the requirement that one of the pair of leptons has a total mass roughly around the nominal mass, 91 GeV, of the Z boson, i.e. between 50 and 106 GeV (ATLAS Collaboration 2012, p. 3).

It is, however, again possible that reactions in which the four leptons are produced in other ways than through the decay of a $ZZ^{(*)}$ pair also pass the selection criteria, see the middle part of Fig. 2. Like misidentification, these reactions are part of the "reducible background" as it is called by the physicists. At this stage of the analysis, the reducible background consists of reactions that end in final states with *almost the*

[6] This point was brought to my attention by Markus Zinser.

[7] For a virtual particle, the usual relation between the rest mass, the energy and the momentum of a particle is not satisfied.

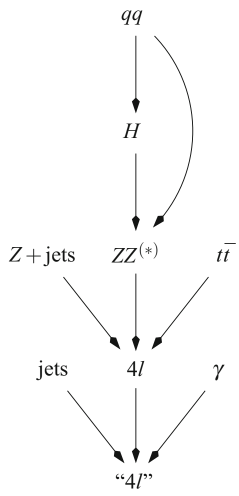


Fig. 2 Causal pathway, including the most important alternative causes, from the newly discovered particle (which may be the Standard Model's Higgs boson) to the reactions in the detector material that are supposedly caused by four leptons resulting from the decay of the Higgs boson. The *arrows* denote the causal relevance relation "is one of the causes for". H denotes the presence of the newly discovered particle, qq a quark pair contained in the colliding protons, Z a Z boson (one of the gauge bosons of the electroweak interaction), $Z^{(*)}$ a possibly virtual Z boson, t a top quark, $\bar{t}$ an anti-top quark, jets are hadronic jets, l denotes the presence of a lepton, γ the presence of a photon (light-quantum)

same characteristics than the sought-for final states from the Higgs decay. It mainly comes from intermediate states consisting of a Z boson plus hadronic jets or a pair of top and anti-top quarks ($t\bar{t}$).

Again the procedure can be interpreted as a diagnostic causal inference where alternative causes are excluded, as much as possible, through data selection. Also, it is again possible, at least in principle, to estimate the average fraction (including its fluctuation) of selected events caused by the remaining alternative causes using only the experimentally well confirmed part of the Standard Model such as its statements about the decay rates of Z bosons, hadronic jets, and $t\bar{t}$ pairs into four leptons. Although the average energies of these reactions at the energies of a hypothetical production of a Higgs boson are higher than the average energies in the known processes, the reactions in question take place at sufficiently high energies sometimes also in the known processes. And even if this is not the case, I think it is fair to say that the general causal relevance relations, as described by the Standard Model, are sufficiently well and systematically established experimentally at lower energies such that the extrapolation to higher energies is as safe as one can hope.

The final step in the justification of the observation of a new particle decaying into four leptons is again, on my account, a diagnostic causal inference. This time the researchers infer the existence of a new particle and one of its most characteristic parameters, its mass. Through data selection the researchers now try to find a certain number of events that is highly unlikely to have been produced by any of the known causes or disturbing factors. If they succeed they can already infer, on the basis of the principle of causality, the likely existence of an unknown cause, which, however, would be completely unspecified.

To infer the slightly more specific claim that a new particle with a certain mass (and integer spin, i.e. a "boson") exists, the researchers can use well established causal relevance relations such as the conservation of energy, momentum and mass (and spin) in particle decays. First of all, the events that are highly unlikely to have been caused by known causes are still instances of a type of event "presence of certain configuration of known objects such as electrons, muons, photons, and hadronic jets". Also, in previous experiments, the causal relevance relation has been sufficiently established that particles with a certain mass will decay, if they decay at all, into other particles and quanta (e.g. photons) such that the relativistic total energy and momentum is conserved. For all we know and can reasonably expect, also the Standard Model Higgs boson, which supposedly is involved in processes of mass generation (via the "Higgs mechanism"), will instantiate such a causal relevance relation. However, the more detailed the specification of the type of cause gets the more alternative causes for the effect have to be taken into account. For instance, the particle involved in the cause may share the properties like mass and spin with the hypothetical Standard Model Higgs boson but may not be involved in processes of mass generation.

The researchers can either rely on more specific regularities and thus make their discovery claim more specific but also more liable to the possibility that alternative causes were at work, or rely on more general regularities and thus arrive only at a more general discovery claim but one that is less liable to the possibility of alternative causes. At this point, the researchers have to strike a pragmatic balance between the two possibilities. In my view, they tend to commit themselves to rather general claims and thus to exclude virtually any serious alternative. If pressed hard, I imagine the researchers would even acknowledge that they have just found "something" with such and such mass, whatever it may be exactly. This is less than one might hope, especially when one is interested in processes of mass generation, but has the distinct advantage of circumventing any serious challenge of unconsidered, or even unconceived, alternatives.

Another problem the researchers have to deal with at this stage is that, even in principle, because of quantum mechanical superposition, it cannot be excluded that some of the selected events are caused by the known particles. Physicists speak of "irreducible background" in such a case. According to the well-confirmed part of the Standard Model, $ZZ^{(*)}$ pairs can be produced directly from the collisions of the quarks contained in the proton beam, see the uppermost part of Fig. 2. But, as with reducible background, the fraction of those events, and the statistical fluctuation of that fraction, can, at least in principle, be estimated using the well-confirmed part of the Standard Model even if the previous experiments only reached lower energies, or produced the reactions in question at sufficiently high energies only at a low rate. Previous experiments on direct production of $ZZ^{(*)}$ pairs have confirmed the Standard Model's description of it, including its dependence on the energy of the reaction, such that an extrapolation to higher energies is warranted.[8]

[8] It is custom among physicists to assume that interference effects in this kind of analysis are negligible (see, e.g., Sjöstrand et al. 2006, p. 10). The assumption seems to be warranted in the present case (see ATLAS Collaboration 2012, p. 5, and reference therein to the preprint of Kauer and Passarino 2012).

In this way, the researchers are able to present a selected sample of events of which it is highly unlikely that the known particles have caused them all. The principle of causality then justifies the inference to the likely existence of an unknown cause for some of those events. The subsumption of the selected events under a type of effect of the well established causal relevance relation, described by the law of energy conservation, further allows the inference of the mass of the unknown cause (presumably a particle). Since the types of events that are used for the diagnostic causal inferences are hypothetical effects of the Standard Model's Higgs boson, the researchers conclude, in addition, that the characteristics of the new particle are compatible with those of the Higgs boson.[9]

4 Summary and discussion

4.1 Causal inference to unobservable entities

I have described an important part of the reasoning from data that justifies the recent discovery of a new particle at the Large Hadron Collider at CERN as a series of diagnostic causal inferences. That is, from the instantiation of certain types of effects, the researchers infer the instantiation of certain types of causes. My main aim was to show how the researchers can infer the existence of an unobservable entity by methods of causal reasoning. The main challenge consisted in showing how the methods of causal reasoning can be employed without presupposing the existence or presence of the entity.

The challenge can be met to a significant extent by realizing, first, that the principle of causality alone allows for the inference to an unknown cause of certain observed effects even without knowing any causal relevance relations; second, that causal relevance relations between relatively general types of causes allow for interesting diagnostic causal inferences. Such general causal relevance relations are reasonably believed to be instantiated by the entity in question, if it turns out to exist, even though the causal relevance relation had to be established without the knowledge of the entity's existence.

On the resulting account, causal reasoning can be used to infer hitherto unknown unobservable entities to the extent to which they are construed as instantiating well-established, albeit rather unspecific, causal relevance relations such as the conservation of mass, charge, and spin.

In conclusion I will sketch some of the likely consequences of my account of the Higgs discovery. They concern interpretational problems of quantum theories, the role of theoretical predictions, the function of data selection, and the problem of unconceived alternatives.

[9] "This observation [...] is compatible with the production and decay of the Standard Model Higgs boson" (ATLAS Collaboration 2012, p. 1).

4.2 Causality in particle physics

The principle of causality plays a key role in my account of the Higgs discovery, both when it comes to the inference of the existence of some unspecified unknown cause as well as when it comes to the more specific claim that there exists an object with such and such mass. However, elementary particle processes fall in the domain of the quantum theory of fields, which shares virtually all interpretational problems of quantum mechanics, in particular the difficulties associated with the notorious violation of John Bell's inequality by empirical data (Bell 1964).

The empirical violation of the inequality forces us to give up at least one of the very basic principles characterizing scientific theories, and the principle of causality is one of them (see, e.g., van Fraassen 1982). If, for my account of the Higgs discovery, I need the principle of causality in a way that it must be valid for Bell inequality violating phenomena, then I will be led to an interpretation of quantum mechanics that gives up one of the other involved principles, such as "locality", i.e. the requirement that causes and effects must be spatio-temporally close to each other.[10] However, this is a route many other authors are willing to take anyway (see, e.g., Maudlin 2011; cf. Wüthrich 2014).

Another possible conclusion to draw from the tension between my account of the Higgs discovery and the interpretational difficulties of quantum mechanics is that elementary particle physics, though concerned with quantum theoretic phenomena, seems to be sufficiently "classical" such that any violation of the principle of causality that may be present in quantum processes does not manifest itself at the level of the reasoning of the experimenters. This could be the case because, for the discovery of new particles through diagnostic causal inferences, temporal sequences ("decay chains") rather than correlations of space-like separated events are the relevant configurations. And since correlations of space-like separated events are typically the only configurations where Bell's inequality or similar constraints are violated, the researchers who are trying to discover a new particle need not typically deal with such processes, in which the principle of causality may not be valid.

4.3 The role of theoretical predictions

To the extent to which the ATLAS Collaboration only infers the existence of some new object consistent with the properties of the Higgs boson as postulated by the Standard Model, that postulate has only a *heuristic* role in the discovery of the new particle. The researchers use this piece of theory to choose the channels and energy ranges in which to search for a new particle. The *evidential* role, however, of this piece of theory is nil if the discovery is interpreted as a diagnostic causal inference based on previously established experimental knowledge. On such an interpretation, the diagnostic causal inferences that the ATLAS collaboration used to justify the discovery of a new particle

[10] Note that the terminology I use here is different from the one often used in the context of quantum field theory where the term "causality" denotes versions of this latter "locality" assumption. Also, as indicated at the beginning of Sect. 2, I distinguish between "causality" and "determinism".

would be equally well warranted had no theoretical prediction of a new particle been put forward.

Such an interpretation, it seems to me, is at least compatible with the way the ATLAS collaboration expresses their results, even now, more than two years after their "observation of a new particle" (ATLAS Collaboration 2012). Although they do refer to "Higgs boson production" in the title of some of their recent articles (ATLAS Collaboration 2013, 2014), passages in the body of the articles hardly suggest more than *consistency* of their measurements with what can be deduced from the Standard Model.[11]

At any rate, the parts of the Standard Model that describe the production and decay of the known particles are relied upon heavily, even on my account, in the justification of the observation of the new particle. They, together with statistical methods, are used, on my account, to estimate how often alternative types of causes, i.e. the "background", produce the selected events and what fluctuations in this number we should expect. Yet, the parts of the Standard Model that are used for this purpose need not be seen as part of a theory which is tested or confirmed by the discovery experiment. Rather I take them to encapsulate, in a systematic way, the empirical knowledge gained in previous experiments and allow, at least in principle, this knowledge to be extrapolated to regions in the relevant parameter space that have not been attained in the previous experiments.

Given the extreme complexity of the analysis procedures of the ATLAS collaboration, it is difficult to determine whether, in fact, their procedures at some point presuppose the theoretical postulate of the Higgs boson's existence, and I consider this an open question. I dare to claim here that, at least in principle, they do not need such a presupposition.

4.4 The function of data selection

The interpretation of the Higgs discovery as a diagnostic causal inference shows why data selection is a necessary and, usually, a legitimate procedure. The selection of data is the means to find a set of events which is very unlikely to have been produced by the known causes. The researchers apply selection criteria on a given initial set of events (see Fig. 1) in order to arrive at a subset of events which do not have the characteristic properties of events caused by the decay of known particles. The selection criteria are chosen on the basis of the knowledge of the characteristic behavior of the known particles, and it is a legitimate goal of the researchers to arrive at a set for which the probability to be caused by the known particles, i.e. the probability given the "null hypothesis", is particularly low.

All care must be taken, however, in the estimation of that probability. In the past, several controversies arose about the question whether this had been done correctly given the respective selection criteria (see, e.g., Staley 2004; Franklin 1998). These

[11] For instance, they still speak of the "newly observed particle" (ATLAS Collaboration 2013, p. 88), and of "the discovery of a new particle" (ATLAS Collaboration 2014, p. 234), rather than of the (Standard Model's) Higgs boson, in the introductions to their papers. See also footnote 9.

cases bring to the fore potentially problematic aspects of data selection, which have to be taken into account in order to perform the necessary exclusion of alternative causes through data selection in a justified and legitimate way.

4.5 The problem of unconceived alternatives

The problem of unconceived alternatives (Stanford 2006) is circumvented, on my account, by interpreting the ATLAS collaboration's discovery claim as of such general nature that it leaves no worthwhile alternatives left. On my reading of the collaboration's publications they do not make more specific claims about the nature of the newly discovered object than what can be concluded from well-established and general causal relevance relations involving conservation of quantities such as energy, momentum, charge and spin. The fit of the experimental findings with more specific theoretical predictions commits the collaboration to hardly more than a consistency claim.

The price to pay, if you wish to call it like that, for this way of dealing with unconceived alternatives is that the experiment did neither test the Standard Model, nor increase its probability, nor otherwise raise its status, over and above the fact that the ATLAS collaboration's findings are compatible with the theory. But such an interpretation of the collaboration's discovery claim seems compatible with their way of expressing themselves and shows how and to what extent the discovery of an unobservable entity can be achieved by causal reasoning.

Acknowledgements Work for this article has been funded by the Swiss National Science Foundation (grant no. 145409), and has been carried out at the Centre for Philosophy of Natural and Social Science of the London School of Economics and Political Science, the Max Planck Institute for the History of Science, and at my home institution, the Technical University Berlin. The central ideas presented here originated in discussions at the history and philosophy of science unit at the University of Bern, in particular within the "MetaATLAS" group consisting of Gerd Graßhoff, Basil Marti, Maya Schefer, and me. I also thank Hans Peter Beck for valuable discussions, and the two reviewers for their detailed and constructive comments.

References

Arabatzis, T. (2012). Hidden entities and experimental practice. In S. Mauskopf & T. Schmalz (Eds.), *Integrating history and philosophy of science: Problems and prospects* (pp. 125–139). Dordrecht: Springer.

ATLAS Collaboration (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, *716*, 1–29.

ATLAS Collaboration (2013). Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC. *Physics Letters B*, *726*, 88–119.

ATLAS Collaboration (2014). Fiducial and differential cross sections of Higgs boson production measured in the four-lepton decay channel in pp collisions at with the ATLAS detector. *Physics Letters B*, *738*, 234–253.

Baumgartner, M. (2008). Regularity theories reassessed. *Philosophia*, *36*, 327–354.

Baumgartner, M. (2009). Uncovering deterministic causal structures: A Boolean approach. *Synthese*, *170*, 71–96.

Baumgartner, M. (2013). A regularity theoretic approach to actual causation. *Erkenntnis*, *78*(1), 85–109.

Baumgartner, M., & Graßhoff, G. (2004). *Kausalität und kausales Schliessen. Eine Einführung mit interaktiven Übungen*. Bern: Bern Studies in the History and Philosophy of Science.

Bell, J. S. (1964). On the Einstein–Podolsky–Rosen paradox. *Physics*, *1*(3), 195–200.
Bohm, D. (1952a). A suggested interpretation of the quantum theory in terms of “hidden” variables. I. *Physical Review*, *85*(2), 166–179.
Bohm, D. (1952b). A suggested interpretation of the quantum theory in terms of “hidden” variables. II. *Physical Review*, *85*(2), 180–193.
Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Clarendon Press.
CMS Collaboration (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, *716*(1), 30–61.
Egg, M. (2012). Causal warrant for realism about particle physics. *Journal for General Philosophy of Science*, *43*(2), 259–280.
Franklin, A. (1998). Selectivity and the production of experimental results. *Archive for History of Exact Sciences*, *53*(5), 399–485.
Graßhoff, G., & May, M. (2001). Causal regularities. In W. Spohn, M. Ledwig, & M. Esfeld (Eds.), *Current issues in causation* (pp. 85–114). Paderborn: Mentis.
Graßhoff, G., & May, M. (2003). Hans Krebs and Kurt Henseleit’s laboratory notebooks and their discovery of the urea cycle: Reconstructed with computer models. In F. Holmes, J. Renn, & H. J. Rheinberger (Eds.), *Reworking the bench: Research notebooks in the history of science* (pp. 269–294). Dordrecht: Kluwer.
Hempel, C. (1966). *Philosophy of natural science*. Englewood Cliffs, NJ: Prentice Hall.
Kauer, N., & Passarino, G. (2012). Inadequacy of zero-width approximation for a light Higgs boson signal. *Journal of High Energy Physics*, *2012*(8), 1–27.
Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). London: Routledge.
Maudlin, T. (2011). *Quantum non-locality and relativity: Metaphysical intimations of modern physics*. Malden: Wiley-Blackwell.
Nickelsen, K., & Graßhoff, G. (2011). In pursuit of formaldehyde: Causally explanatory models and falsification. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *42*(3), 297–305.
Rappaport, S. (1996). Inference to the best explanation: Is it really different from Mill’s methods? *Philosophy of Science*, *63*, 65–80.
Scholl, R. (2015). Inference to the best explanation in the catch-22: How much autonomy for Mill’s method of difference? *European Journal for Philosophy of Science*, *5*(1), 89–110.
Sjöstrand, T., Mrenna, S., & Skands, P. (2006). PYTHIA 6.4 physics and manual. *Journal of High Energy Physics*, *2006*(05), Article 026.
Spirtes, P., Glymour, C. N., & Scheines, R. (1993). *Causation, prediction, and search*. New York: Springer.
Staley, K. (2004). *The evidence for the top quark: Objectivity and bias in collaborative experimentation*. Cambridge: Cambridge University Press.
Stanford, P. K. (2006). *Exceeding our grasp: Science, history, and the problem of unconceived alternatives*. New York: Oxford University Press.
Suárez, M. (2008). Experimental realism reconsidered: How inference to the most likely cause might be sound. In S. Hartmann, C. Hoefer, & L. Bovens (Eds.), *Nancy Cartwright’s philosophy of science* (pp. 137–163). New York: Routledge.
van Fraassen, B. C. (1982). The Charybdis of realism: Epistemological implications of Bell’s inequality. *Synthese*, *52*, 25–38.
van Fraassen, B. C. (1989). *Laws and symmetry*. Oxford: Clarendon.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.
Wüthrich, A. (2012). Methoden des Nachweises von Elementarteilchen: Die (Wieder-)Entdeckung des W-Bosons 1983 und 2010. In G. Graßhoff, & A. Wüthrich (Eds.), *MetaATLAS: Studien zur Generierung, Validierung und Kommunikation von Wissen in einer modernen Forschungskollaboration* (pp. 215–264). Bern: Bern Studies in the History and Philosophy of Science.
Wüthrich, A. (2014). Local acausality. *Foundations of Physics*, *44*(6), 594–609.