

Causal Warrant for Realism about Particle Physics

Matthias Egg

Department of Philosophy, University of Lausanne

CH-1015 Lausanne, Switzerland

e-mail: matthias.egg@unil.ch

August 20, 2012

Abstract While scientific realism generally assumes that successful scientific explanations yield information about reality, realists also have to admit that not all information acquired in this way is equally well warranted. Some versions of scientific realism do this by saying that explanatory posits with which we have established some kind of causal contact are better warranted than those that merely appear in theoretical hypotheses. I first explicate this distinction by considering some general criteria that permit us to distinguish causal warrant from theoretical warrant. I then apply these criteria to a specific case from particle physics, claiming that scientific realism has to incorporate the distinction between causal and theoretical warrant if it is to be an adequate stance in the philosophy of particle physics.

Keywords Causal explanation · Entity realism · Inference to the best explanation · Neutrino · Particle physics · Scientific realism

1 Introduction

The debate on scientific realism has taken various directions in the past few decades, but one fairly pervasive feature has been the growing awareness that realism should not be viewed as an all-or-nothing affair. Even realists like Stathis Psillos, who aim to provide general arguments for a far-reaching scientific realism, give statements like the following:

If scientific realism is to be plausible and, as most realists would urge, in agreement with actual scientific practice, then it must go for differentiated commitments to scientific theories, and what they entail about the world, in accordance with the evidence which supports them, as a whole and in parts. (Psillos 1999, 161)

The interesting question is how such a differentiation of commitment is to be attained. Unfortunately, realists have often been so busy fighting off antirealist criticism that they failed to address this question with sufficient precision. In particular, realists like Psillos argue for the validity of a general inference scheme, known as *inference to the best explanation* (IBE), without asking whether all instances of this scheme are equally well warranted.

This article seeks to improve on that situation by distinguishing two types of warrant (*theoretical* and *causal*) that can be generated by IBE. The origin of this distinction can be traced back to Nancy Cartwright's (1983) version of *entity realism* (or *experimental realism*), and it was made explicit by Mauricio Suárez (2008). The basic idea is that causal warrant is stronger than theoretical warrant and that accordingly, scientific claims which are causally warranted are worthy of a particularly strong commitment on the realist's part. I will use the term *causal realism* for a realism that incorporates this idea.

The central distinction between causal and theoretical warrant will be introduced in section 2 by providing three (individually necessary and jointly sufficient) criteria for causal warrant. As Suárez's paper has already discussed some aspects of this distinction, my discussion will focus on some aspects which he has neglected. Section 3 will then show that a causal realism based on this distinction does no longer fall prey to the objections that were raised against Cartwright's entity realism. In section 4, I will apply these general considerations to a concrete case from the history of particle physics, the discovery of the neutrino. The importance of causal warrant will be highlighted by the fact that any position failing to appreciate this concept will be unable to grasp the significance of what physicists call "the detection of the neutrino". Section 5 contains some concluding remarks on the strength of causal warrant in the face of antirealist criticism.

2 Distinguishing Causal from Theoretical Warrant

Cartwright's (1983) version of entity realism rests on the distinction between theoretical and causal explanation and on the claim that IBE works reliably only in the latter, not in the former case. Or, to adopt the terminology from Suárez (2008), *inference to the most likely cause* (IMLC) is a valid inference scheme, while *inference to the best theoretical explanation* (IBTE) is not. But as we will see in section 3, there are good reasons for thinking that the difference between causal and theoretical explanation is not as clear-cut as Cartwright originally envisaged. This is the starting point of Suárez's (2008) paper.

Like Cartwright, Suárez holds that causal explanation provides stronger grounds for existential commitments than theoretical explanation. But instead of drawing a sharp contrast between IMLC as a valid inference scheme on one side and IBTE as an invalid one on the other side, Suárez construes the difference between the two types of inference as a difference between two types of warrant that are generated by IMLC and IBTE, respectively. He thereby weakens Cartwright's (1983, 93) claim that accepting a causal explanation provides "conclusive reason" to believe in the reality of the cause:

'Conclusive' is thus to be understood as a relative term: Causal warrant is

conclusive in comparison with the warrant provided by an IBTE. No existential commitment derived from an IMLC can be defeated by any amount of theoretical warrant to the contrary. [...] The only defeater of causal warrant in favour of the existence of x is causal warrant of the same strength against x . (Suárez 2008, 145)

But even with this relativization in place, the causal realist faces the challenge of explicating what it is that differentiates causal from theoretical warrant. As I see it, Suárez's reply to that challenge is on the right track, but incomplete. In the following presentation of the distinction between causal and theoretical warrant, I will therefore be quite brief on those aspects which I think Suárez already got right, while elaborating on what needs to be added to his characterization in order for it to be completely convincing.

In a nutshell, I characterize theoretical and causal warrant as follows:

1. Every instance of IBE generates *theoretical warrant*.
2. An instance of IBE generates *causal warrant* if and only if the corresponding explanation additionally fulfills the criteria of non-redundancy, material inference and empirical adequacy.

I now turn to a discussion of these three criteria.

2.1 Non-redundancy

If we are to infer the truth of a hypothesis on the basis of its ability to explain a given phenomenon, we need to make sure that no other hypothesis provides an equally satisfactory explanation of the phenomenon. Concerning the way in which such redundancy is removed, Suárez identifies an important difference between causal and theoretical explanations:

The nonredundancy requirement is met to a much larger degree by IMLC than by IBTE. Scientists establish which putative cause is nonredundant through controlled intervention and manipulation in laboratory conditions—and only then have they got reason to believe that the putative cause is genuinely responsible for the phenomenon.

By contrast, we have much poorer and controversial methods to establish which among all possible empirically adequate theoretical explanations of a phenomenon is the most probable one. [...] Is the most probable theoretical explanation the simplest one, the most ontologically parsimonious, the most familiar, the one that preserves the greatest amount of structure from previous theories, the one that explains a greater number of independent phenomena, etc, etc, etc. As a consequence there is much more redundancy, in the form of underdetermination, in the case of theoretical explanation. (Suárez 2008, 155)

If this is correct, then it is reasonable to spell out the criterion of non-redundancy in the following way: A hypothesis is non-redundant exactly if there is no other hypothesis that agrees with the experimental results. Accordingly, I will not count an explanation as non-redundant if it is merely marked out by theoretical virtues of the type mentioned in the penultimate sentence of the above quotation.¹

However, some objections can be raised against the claim that there is much more redundancy in theoretical than in causal explanations. Let us first look at Margaret Morrison's (1994, 127) assertion that "causal explanation exhibits the same kind of redundancy present in theoretical explanation". Morrison's paper is directed against Cartwright (1983, 78–81), who uses an example from laser theory to illustrate the contrast between redundant theoretical and non-redundant causal accounts. Discussing the very same case, Morrison concludes that there is actually no such contrast. It seems to me, however, that what is established by Morrison's analysis is not so much a *redundancy* of causal accounts but a *multiplicity*. And multiplicity of causal accounts by itself is no threat to the validity of IMLC. Instead, it is what we should expect, especially if we insist (as Morrison rightly does) that causes need to be specified more precisely than in Cartwright's generic causal story. It should not surprise us that a phenomenon can in general not be attributed to a single cause, but usually results from a combination of several causal factors or partial causes. It is in this sense that we may tell more than one causal story, but there is no redundancy here, because all these stories are needed to get the causal account right. Of course, we need to specify in each case precisely which causal factors are relevant (and to what degree), but this is in general possible by the same methods we have already encountered in the above quotation: controlled intervention and manipulation in laboratory conditions.² But now Morrison can argue that this need for specification again threatens the distinction between causal and theoretical explanations:

Just as the mathematical treatment is determined by the kind of phenomena one is dealing with, causal stories also involve specific contextual features rather than a generic story applicable to all cases. A generic story could not qualify as a phenomenological law since it would require a *ceteris paribus* clause to guarantee its truth. Instead, the phenomenological laws are the ones that govern each individual case, but if we acknowledge their uniqueness as a basis for their 'truth' then we must also acknowledge the uniqueness of specific theoretical descriptions. (Morrison 1994, 136)

It is debatable in how far the phenomena really *determine* the theoretical treatment in the same way as they determine which causal stories have to be taken into account, but

¹This might be a slightly stronger understanding of non-redundancy than what Suárez has in mind. He seems to trace the redundancy in the case of theoretical explanation to the fact that there are so many different (and possibly conflicting) theoretical virtues. Thus, it seems that he would count an explanation as non-redundant if it is favored by all (or most) theoretical virtues, even if there is an empirically adequate (but unambiguously less virtuous) competitor. By contrast, such an explanation would count as redundant on my understanding of the criterion. In subsection 4.2, I will illustrate and justify this stronger reading of the criterion in the context of the neutrino hypothesis.

²For a well-developed account of causation along these lines, see Woodward (2003).

I grant that uniqueness of theoretical description is sometimes achievable. The lesson to be drawn from this is that non-redundancy is a necessary but not a sufficient condition for causal warrant. Non-redundancy of an explanation by itself does not yet guarantee that the explanatory hypothesis is causally warranted, other criteria need to be taken into account as well.

The difference between theoretical and causal warrant in terms of redundancy can also be questioned on another level, in connection with the general discussion of the *underdetermination of theories by evidence* (UTE).³ One might think that there is always (that is, also in the case of causal explanations) a multitude of mutually incompatible possible explanations between which experimental evidence of the type Suárez speaks of is unable to decide. Consider, for example, a hypothesis H which constitutes a genuine (causal or theoretical) explanation for some phenomenon. Now it is always possible to construct an alternative hypothesis H' asserting that the world behaves exactly as H says if the universe is observed, but if no one is looking, the world behaves in ways which are incompatible with H .⁴ Obviously, we should not let cases like this cast doubt on our warrant for belief in H , even though there is (by construction) no empirical way to determine whether H or H' is true. The reason is simply that this kind of UTE does not just threaten our warrant for scientific hypotheses but for virtually *all* our knowledge. It is, of course, possible to deny that any claim can be warranted, but to do so is to leave the debate on *scientific* realism and to go for general skepticism. Thus, there is a sense in which also the causal realist relies on the evaluation of theoretical virtues: If a hypothesis is so *ad hoc* that taking it into account would result in general skepticism, then it is excluded even if there is no experimental evidence against it.

A more interesting question about the status of the theoretical virtues arises in cases of *non-trivial* underdetermination (or redundancy), that is, cases where mutually incompatible, scientifically respectable hypotheses account for the same phenomena. Here the causal realist denies that the evaluation of theoretical virtues can effectively eliminate redundancy. This attitude resembles what Psillos calls the “ignorance response” to UTE, which consists in admitting that, if more than one theory is compatible with some evidence, we simply do not know (and may never be in a position to know) which of the competing theories is true. Psillos thinks that this concedes too much ground to the antirealist:

³UTE is not quite the same phenomenon as redundancy of explanations, but the difference is irrelevant here. Nevertheless, it may be helpful to spell out how the two concepts differ. Apart from the fact that UTE is not restricted to explanatory contexts, the main difference between UTE and redundancy of explanations is, I think, this: In cases of UTE, the theories in question are generally required to be empirically adequate, whereas (as we shall see in subsection 2.3) no such requirement is made in the case of redundant (theoretical) explanations. This accounts for the rather puzzling fact that no one (to my knowledge) has so far contested Cartwright’s (1983, 79) claim that “this kind of redundancy of treatment [...] is common throughout physics”, while it is generally agreed that cases of genuine (non-trivial) UTE are very rare (Stanford 2001, S5–S7).

⁴The example is modeled on a similar example from Kukla (1993). This and other strategies pursued in the literature to construct empirically equivalent rivals of existing theories are concisely reviewed in Stanford (2001, S2–S7), which is also the source of my argument against taking these cases seriously in the context of scientific realism.

But admitting that the ‘ignorance response’ is a viable realist answer to UTE is tantamount to conceding the main point that realists need to defend: that there is space for rational belief in one of the two empirically congruent theories. Instead of addressing the real challenge that UTE poses to realism—that there is no possibility of warranted belief in one of the two theories—the ‘ignorance response’ simply sidesteps it. (Psillos 1999, 169)

This criticism has some force to it, so it would be unwise for the causal realist to completely reject the epistemic value of the theoretical virtues. On the other hand, Suárez is right that comparing hypotheses by experimental tests is much more reliable than evaluating their theoretical virtues. Requiring non-redundancy (in the sense defined here) as a necessary criterion for causal warrant takes into account both of these acknowledgements. The theoretical virtues have their role to play in the generation of theoretical warrant. However, they are incapable of generating causal warrant, because they cannot eliminate redundancy. For this, experimental testing is needed.

2.2 Material Inference

An important difference between theoretical and causal explanations has to do with the kinds of entities that appear in the explanans. While the key ingredients for theoretical explanations are abstract entities, namely the laws from which the explanandum is derived,⁵ the explanatory role in causal explanations is typically played by concrete particulars (objects, events etc.), as when, for example, someone finds a pebble in his shoe and says: “Now, *that* explains the soreness of my little toe!”

Suárez (2008, 157–159) describes this difference by saying that IBTE is a *formal inference*, whereas IMLC can be carried out as a *material inference*. Then, comparing two examples of inference, one in the formal, the other in the material mode, he concludes that “the latter type of inference is more robust—it contains fewer steps where it may go wrong” (158). While I agree with Suárez’s conclusion, I find his argument in support of it far from convincing. To see how it could be improved, let us look at Cartwright’s argument for the difference in robustness between IMLC and IBTE:

[C]auses make their effects happen. We begin with a phenomenon which, relative to our other general beliefs, we think would not occur unless something peculiar brought it about. [...] the peculiar features of the effect depend on the particular nature of the cause, so that—in so far as we think we have got it right—we are entitled to infer the character of the cause from the character of the effect.

⁵ The idea that the explanandum is logically derived from the explanans obviously stems from the DN-model of explanation, and a referee for this journal has pointed out that the advocate of IBTE should not be saddled with this outmoded account of explanation. However, the most appropriate candidate for a more modern approach to theoretical explanation is the unification model developed by Michael Friedman and Philip Kitcher, and this approach “retains the Hempelian idea that to explain a phenomenon is to produce an argument whose conclusion describes the phenomenon” (Kitcher 1989, 431).

But equations do not bring about the phenomenological laws we derive from them (even if the phenomenological laws are themselves equations). Nor are they used in physics as if they did. (Cartwright 1983, 76)

According to Cartwright, the difference between causal and theoretical explanations lies in the fact that the explanans “brings about” the explanandum only in the former, not in the latter case. In order to grasp how this relates to the difference between material and formal inference, we need to understand more precisely what “bringing about” means. I suggest the following explication: If x brings about y , then (insofar as y occurs at all), there are some tokens of y which would not have occurred if x had not occurred.⁶ The truth of this counterfactual statement for at least some tokens of y seems to be an essential part of what it means for x to cause y (regardless of the difficulties that a general account of causation in terms of counterfactuals may face), and, in conjunction with a sufficient number of y -occurrences, it implies realism with respect to x .

Now it might seem that there is a similar counterfactual dependence between the explanandum and the explanans in the case of a theoretical explanation, because such explanations contain laws, and an essential aspect of lawhood is the ability to support counterfactuals. But these counterfactuals are not of the right sort to support realism about laws in the same way as the above counterfactuals support realism about causes. Here is why: Laws support counterfactual claims concerning their instances. For example, Boyle’s law supports claims like “if the volume of this gas was reduced at constant temperature, its pressure would increase”. But what IBTE aspires to establish is not the truth of a singular statement but of the law itself. In order to achieve this, a claim of the following form would be needed: “If law L did not hold, phenomenon y would not occur.”⁷ This is actually not just a counterfactual, but a counterlegal statement. There may be ways to evaluate such statements, but it is clearly not part of scientific explanation to do so. By contrast, as we have seen above, the counterfactual statements that give rise to causal realism are an integral part of causal explanations.

Therefore, the connection between scientific explanation and realism is much tighter in the case of causal explanations than in the case of theoretical explanations. It is this difference that the criterion of material inference aims to capture by stating that there is causal warrant only if the result of an IBE is a concrete matter of fact, as opposed to a law.

The soundness of this characterization stands and falls with how one draws the line between concrete matters of fact and laws. Does, for example, the statement that electrons are negatively charged express a concrete matter of fact (because electrons are concrete entities) or a law (because it is a general statement)? If the latter, then it seems that hardly any statement of scientific interest can be arrived at by material inference, so that a realism based on this criterion faces the threat of vacuity. (As we will see in subsection 3.1, this was one of the main objections against entity realism.) On the other hand, if the

⁶The restriction to *some tokens* accommodates the fact that y could also come about in other ways.

⁷By speaking of y as an occurrent phenomenon, I am assuming that the explanandum is a singular fact, but the argument is just the same if y is itself a (phenomenological) law, as Cartwright assumes in the above quotation.

negative charge of electrons counts as a concrete matter of fact, why should the fact that electrons are accurately described by quantum electrodynamics (QED) not also count as such a fact? But then it seems that the criterion of material inference would be trivially satisfied by even the most theoretical explanations, if only they contain references to at least *some* concrete entities.

To solve this dilemma, we need to take a closer look at the causal claims which underlie IMLC. And this firstly requires some clarification on the relata of the causal relation. So far, I have been very permissive in this respect, allowing objects, events, facts or phenomena to figure as causes or effects. My reason for doing so is that I do not think that any of these is *the* right way to speak about causation. They are all useful in certain contexts (and I will continue to use them whenever there is no danger of confusion), but if we want to be precise, we have to speak of causation in terms of causal *properties* (see Chakravartty 2007, 107–111). This is not the place to enter into a detailed metaphysical discussion of causal relata, but the observation just made helps us to see more clearly what needs to be done in order to flesh out the criterion of material inference: What is to count as material inference needs to be defined in terms of the kinds of properties that can be ascribed to entities as a result of such inference (let us call them *material properties*).

Which properties belong to that class? An obvious proposal would be to identify them with Anjan Chakravartty’s (2007, 41) *causal properties*, that is, properties which “confer dispositions for relations, and thus dispositions for behaviour on the particulars that have them”. But this is not precise enough, because the above-mentioned property of “being accurately described by QED” also confers a disposition for behavior on the electron (at least in some sense).⁸ The necessary refinement of this proposal can be found, I think, in James Woodward’s manipulability account of causation. Woodward (2003, 112) holds that “for something to be a cause we must be able to say what it would be like to change or manipulate it”.⁹ The rationale for this requirement is that we usually test causal claims by intervening on the putative cause and observing whether the putative effect changes accordingly. While it is, on Woodward’s account, not essential for causal claims that such an intervention is possible in practice (see Woodward 2003, 86–91, 127–133), it is essential that there is a well-defined notion of what it would mean to intervene on the putative cause. This requirement furnishes us with the necessary distinction between material and formal properties. Let us apply it to the two examples constituting the above dilemma. It is perfectly clear what it means to modify the property of negative

⁸On the other hand, it would be too restrictive to identify the material properties with what Chakravartty (2007, 47) calls *detection properties*, that is, “causal properties one has managed to detect”. The criterion of material inference should only be concerned with what *can* be established, not with what *has been* established. Otherwise, the criterion would by itself be sufficient for causal warrant, thus rendering the other criteria superfluous. So what we need here is not a specification of *detection properties*, but of *detectable properties*.

⁹With regard to the question of what the relata of the causal relation are, Woodward introduces yet another candidate, namely *changes in the values of variables*. He also mentions some cases in which thinking of causation in terms of properties (as Chakravartty does) would lead us astray (Woodward 2003, 112–114). However, Chakravartty would probably deny that the properties appearing in these examples are *causal properties*, so the conflict between the two approaches seems to be resolvable.

charge. (In some cases, this is even practically possible, by performing an experiment with positrons instead of electrons, but we may also envisage hypothetical experiments in which the charge of the electron is set to a value other than $\pm e$.) By contrast, it is not clear what it would mean to intervene on the property of being accurately described by QED. It could mean to modify any of the various processes that QED describes, or to introduce an entirely new interaction, etc. So we arrive at the following definition of material inference: an inference in the material mode is one that results in ascribing to a concrete entity a property for which there is a well-defined notion of what it means to modify it.¹⁰

To conclude the discussion of material inference, I must address a more fundamental objection to the legitimacy of this criterion. Many will view the causal notions which appeared throughout the discussion as an element of dubious metaphysics. They deny that there are causes in nature and that advanced science is concerned with causal claims. One way to answer this objection would be to enter into a discussion of the fundamental structure of the world and to argue that causation has a role to play there. There are well-developed arguments for this, both from pure metaphysics (e.g., Bird 2007) and from the reflection on contemporary physics (Esfeld 2010); furthermore, Chakravartty (2007) has recently integrated a metaphysics of causal properties into a defence of scientific realism.

While I am not opposed to this line of reasoning, I suspect that it claims more than what the realist needs. What these authors defend is a kind of *causal fundamentalism* (the view that causality is a fundamental trait of nature), while *causal realism*, as I understand it, can be argued for independently of how one thinks about these fundamental issues. For this move, I take inspiration from one of today's fiercest critics of causal notions, John Norton. In his essay *Causation as folk science* (2007), Norton denounces causal fundamentalism, but at the same time admits that causes are in some sense real. To illustrate this, he draws an analogy between causes and gravitational forces: although gravitational forces do not appear in our most fundamental theory of gravitation (general relativity), they have a *derivative reality* in the sense that the classical Newtonian theory incorporating them can be recovered as a limiting case of general relativity. Just like gravitational forces, causes can be considered real, even though they are not fundamental:

[C]auses, caloric¹¹ and gravitational forces have a derivative reality. They are not fictions insofar as they are not freely invented by us. Our deeper sciences

¹⁰While this excludes formal and higher order properties, it is *not* part of causal realism to insist that only *intrinsic* properties are allowed. So there is no contradiction between causal realism and structural realism, as long as the latter conceives of structures as (first-order) relations between particulars. (See Chakravartty 2007, sections 2.2 and 2.3)

¹¹Besides the analogy with gravitational forces, Norton also compares causes to the *caloric*, the alleged heat substance of pre-modern theories of heat. But this analogy is ill chosen, because the caloric theory of heat is not just *superseded* (like Newtonian mechanics) but thoroughly *discredited*. Accordingly, no one would nowadays seriously claim that caloric is real, while many of us hold that gravitational forces are real. Norton's mistake is that he focuses on just one aspect of the caloric theory, namely the claim that heat is a conserved fluid. The actual theory made various other claims about the caloric, and these render a recovery of the caloric theory as a limiting case of today's kinetic theory impossible (see Chang 2003).

must have quite particular properties so that these entities are generated in the reduction relation. [...] But then they cannot claim the same reality as the fundamental ontology. (Norton 2007, 31)

The distinction between the real and the fundamental is of prime importance for the debate on scientific realism, since this debate is based on assuming the reality of objects that are undoubtedly non-fundamental, namely the observable objects of our everyday experience. For the present discussion, the importance of this distinction lies in the fact that it opens the way to treating causal relations as something that really exists in the world, without burdening oneself with too much metaphysical baggage. The criterion of material inference thereby gains a basis that can be accepted by a wider audience than just the supporters of causal fundamentalism.

2.3 Empirical Adequacy

According to Bas van Fraassen (1980, 12), “a theory is empirically adequate exactly if what it says about the observable things and events in this world, is true—exactly if it ‘saves the phenomena’.”¹² Being thus defined as a property of *theories*, empirical adequacy does not seem to have much to do with *causal* warrant. I nevertheless claim it to be a criterion for causal warrant, so I must first explain what it means for a causal explanation to be empirically adequate, and then argue that this condition is typically fulfilled for causal, but not for theoretical explanations.

What an explanation says about observables is not given by the explanatory hypotheses alone but by what can be deduced from them (with the help of some auxiliary hypotheses). Therefore, to speak of the empirical adequacy of an explanation requires that it makes sense to speak of the deductive consequences of the explanans. Now if the explanans of a causal explanation is a concrete particular (as I suggested at the beginning of the last subsection), then it may sound like a category mistake to say that we deduce anything from it. However, the subsequent discussion made it clear that causal explanations are associated with certain descriptions (statements of what I have called concrete matters of fact), and these do, of course, have logical implications. Some of these implications are statements about observable things and events, and if these statements are true, the criterion of empirical adequacy is fulfilled.

But is empirical adequacy not something we should require from *any* explanation, theoretical or causal? After all, even an antirealist like van Fraassen, who denies that explanation is connected with truth, agrees that it is connected with empirical adequacy (this point will be important in subsection 3.3). Nonetheless, Cartwright argues against this connection in the case of theoretical explanations:

[Van Fraassen and realists like Sellars] have in common a surprising respect for theory. Both expect that theories will get the facts right about the observable

¹²Van Fraassen calls this a “preliminary explication” of empirical adequacy. His more complete explication (van Fraassen 1980, chap. 3) relies heavily on the semantic approach to scientific theories, which is not relevant for the present discussion.

phenomena around us. [...] This is not how I see good theories working. The observational consequences may be a rough match to what we suppose to be true, but they are generally not the best we can do. If we aim for descriptive adequacy, and do not care about the tidy organization of phenomena, we can write better phenomenological laws than those a theory can produce. (Cartwright 1983, 159–160)

To support her claim, Cartwright (1983, essay 6) discusses various theoretical explanations in which the explanandum does not follow from the explanans by strict deduction, but only with the help of approximation procedures that improve on the descriptive accuracy of the fundamental laws and are dictated solely by the facts to be derived. In cases like these, it is hard to see how the explanatory relation should provide warrant for the truth of the explanans.

Realists may hope to overcome this predicament by building upon Laudan’s and Leplin’s (1991, 461–465) argument that a theory can be supported by empirical evidence that is not part of its deductive consequences. For the case of IBTE, this would mean that a theory’s explaining of some observed phenomenon would provide warrant for its truth, although the phenomenon is not strictly deduced from it. But I contend that even if this argument is successful, the warrant it provides will clearly be theoretical, not causal. For the “indirect empirical support” that a hypothesis H_1 can, according to Laudan and Leplin (1991, 464), gain from some empirical evidence e that is not directly entailed by it, depends on H_1 being derivable from a more general theory T , which also entails another hypothesis H_2 , of which e is a consequence.¹³ So whatever warrant there is for H_1 derives from warrant for T , and this is only theoretical warrant, because the inference from e to T (via H_2) will in general not satisfy the criterion of material inference. Neither is T ’s ability to unify H_1 and H_2 (or any other theoretical virtue T may possess) sufficient to ensure non-redundancy.

In fact, as seen in the above quotation, Cartwright even argues that a theory’s ability to achieve a “tidy organization of phenomena” works counter to its empirical adequacy. This clarifies why empirical adequacy is more tightly connected to causal than to theoretical explanations. In theoretical explanation, unification of diverse phenomena is essential (see note 5 above); without it, a hypothesis would not count as explanatory. By contrast, though unification is certainly desirable in causal explanations, a causal hypothesis can be explanatory without it, by relying on concrete matters of fact and fairly specific phenomenological laws. Having a more limited range of application, statements about such facts and laws have a higher chance of being empirically adequate.

However, these considerations point to an important ambiguity in the notion of empirical adequacy, which needs to be removed if this criterion is to be coherently applied. Having explicated empirical adequacy as a theory’s ability to “save the phenomena”, van Fraassen (1980, 12) emphasizes that “this refers to *all* the phenomena: these are not exhausted by those actually observed, nor even by those observed at some time, whether

¹³Laudan and Leplin also consider other modes of nonconsequential empirical support, such as analogical reasoning, but these seem even more vulnerable to the charge that what is transmitted here is just theoretical and not causal warrant.

past, present, or future.” A strong reading of “all” in this quote would imply that only a *theory of everything* can be considered as empirically adequate. This would render the notion of empirical adequacy quite useless for our purposes (and, I suppose, for the debate in general too). “All phenomena” must therefore be understood relative to some domain of application. But if we interpret it too liberally, the notion will be equally deprived of its usefulness, for then, empirical adequacy is too easy to get. I illustrate this with a well-known example, which will be discussed in more detail in the next section. In many experiments, electrons behave like tiny billiard balls, moving on well-defined trajectories, subject to local interactions. Still we would not say that this classical picture of the electron is empirically adequate, because there are *other* experiments (e.g., the double slit), producing phenomena which are incompatible with the classical theory. What is needed, therefore, is a specification of a domain neither too vast nor too narrow, within which all the phenomena need to be saved. Furthermore, in order to be suitable for realism, such specifications should be objective in the sense of not depending on the kind of experiment we choose to perform. The most common way to accomplish this is by considering limiting cases with respect to some physical parameter. The example of gravitational forces, familiar from subsection 2.2, illustrates this well. Newton’s theory of gravitation does not save *all* the phenomena, nor even all gravitational phenomena (the perihelion precession of Mercury being a famous counterexample). But its domain of application can be objectively circumscribed by requiring that the considered distances are large compared to the Schwarzschild radii of the considered objects.

3 Causal Realism’s Advantages over Entity Realism

Many of the arguments in the previous section made use of Cartwright’s distinction between causal and theoretical explanation. Given this dependency of my causal realism on (a central element of) Cartwright’s entity realism, I must now show that the points of criticism that were raised against entity realism are no threat to causal realism. I will discuss three charges against entity realism: that it is incoherent, that it is based on a mere convention and that it is contradicted by examples from actual science.

3.1 Incoherence

The entity realist’s mistrust of IBTE is sometimes taken to imply that such a realism is only committed to the *existence* of certain entities but not to the truth of any further statement about them, because that would already require an inference to the best *theoretical* explanation. Such a crude entity realism invites an obvious line of criticism, which is most biting expressed by Alan Musgrave:

To believe in an entity, while believing nothing further about that entity, is to believe nothing. I tell you that I believe in hobgoblins [...] So, you reply, you think there are little people who creep into houses at night to do the housework. Oh no, say I, I do not believe that hobgoblins do that. Actually,

I have no beliefs at all about what hobgoblins do or what they are like. I just believe in them. (Musgrave 1996, 20)

Phrased in this way, the charge is overstated, since no entity realist has ever urged us “to believe in an entity, while believing nothing further about that entity”.¹⁴ But even the coherence of less crude forms of entity realism can be called into question, for if entity realists accept some of the claims about theoretical entities as true, how is this compatible with their not being committed to the truth of the theories that best describe these entities?¹⁵

Causal realism has a clear answer to this question, which was already discussed in subsection 2.2. Theories ascribe different properties to entities, some of them formal, some of them material. Statements concerning the former are based on IBTE, so the causal realist is not (or at most tentatively) committed to them. By contrast, statements concerning material properties can be warranted by IMLC, and if that is the case, belief in these statements does not imply belief in the rest of the theory, for which there is only a weaker kind of warrant.

3.2 Conventionality

Robert Pierson and Richard Reiner (2008) argue that the alleged difference between IBTE and IMLC is largely conventional: due to the work of antirealists like Duhem and van Fraassen, we have grown accustomed to the thought that theories can be explanatory without being true, whereas we may still think that no causal story can be explanatory without citing the actual causes. Another way of expressing this difference is to say that “causally explains” is usually taken to be a *success term*, while “(theoretically) explains” is not (Pierson and Reiner 2008, 272).

Pierson and Reiner question this convention as arbitrary, but this is only the first part of their criticism. They then go on to claim that even if we accept the convention, IMLC will be unable to transmit warrant from its premises (“ P causally explains Q ” and “ Q is true”) to its conclusion (“ P is true”), because if “causally explains” is interpreted as a success term, then the first premise is only warranted insofar as there already is warrant for the conclusion. If P is not describing an actual cause, there is no causal explanation in the first place (278, 281).

However, the criteria discussed in section 2 show that the difference in validity between IMLC and IBTE rests on more than mere convention and that evidence for “ P causally

¹⁴This is obvious in the case of Cartwright and Ian Hacking, who readily profess their belief in “causal principles” and “phenomenological laws” (Cartwright 1983, 8) or “home truths” and “well-understood causal properties” (Hacking 1983, 265). Things are somewhat less clear, however, concerning Michael Devitt (1997), who, for tactical reasons, chooses to defend an entity realism which is merely committed to the *existence* of entities of certain kinds but not to the *properties* science ascribes to them. But when it comes to scientific explanation, even Devitt has to acknowledge that “phenomena are explained not by the mere existence of, say, electrons, but by electrons being the way our theory says they are” (109).

¹⁵This question lies at the heart of the criticisms voiced by Psillos (1999, 256–258; 2008, 169–171) and Chakravartty (2007, 31–32).

explains Q ” can be gathered independently of evidence for the truth of P .¹⁶ This independence is underlined by the fact that each of the three criteria is, by itself, insufficient to guarantee causal warrant. Only when they are jointly fulfilled is there causal warrant for the truth of P .

In this context, another advantage of causal realism over entity realism deserves to be mentioned. Within Cartwright’s entity realism, the criterion of non-redundancy is a necessary condition *for an explanation* to qualify as causal, which has the following consequence: If scientists, after a process of careful experimental testing, finally succeed in eliminating redundancy of explanations for a certain phenomenon, then the one explanation that emerges as the only survivor, having hitherto been a theoretical explanation, will now suddenly become causal. This contradicts the intuition that being causal should be an intrinsic and time-independent feature of an explanation. Cartwright seems to recognize this problem and consequently holds that the causal/theoretical distinction does not depend on whether there *actually is* a redundancy of explanations, but on whether or not *we tolerate* redundancy:

Different [theoretical] approaches are useful for different purposes; they complement rather than compete. [...] We do not have the same pragmatic tolerance of causal alternatives. We do not use first one causal story in explanation, then another, depending on the ease of calculation, or whatever. (Cartwright 1983, 81)

Unfortunately, this way of interpreting the requirement of non-redundancy does little to defend entity realism against the conventionality objection, because it makes the causal/theoretical distinction depend on *our attitude* towards redundancy, which is just as much a matter of convention as our habit of taking “causally explains” (but not “theoretically explains”) as a success term.

But for the causal realist who has assimilated Suárez’s refinement of the causal/theoretical distinction, there is no need to move towards such a subjective interpretation of non-redundancy. If the distinction is no longer between causal and theoretical *explanations* but between causal and theoretical *warrant*, then the above scenario of something theoretical that suddenly becomes causal poses no threat at all. Nobody expects the character of warrant to be intrinsic and time-independent. In the course of inquiry, it repeatedly happens that a theoretically warranted hypothesis at some point acquires causal warrant, and the elimination of redundancy is the obvious way to achieve this transition.

3.3 Counterexamples to the Validity of IMLC

Christopher Hitchcock (1992, sec. IV) argues against the validity of IMLC by giving two counterexamples to the claim that acceptance of a causal explanation in science necessarily involves belief in its truth. This third objection is obviously related to the second one in that it also targets the distinction between IMLC and IBTE. More precisely,

¹⁶For more details on this point, see Suárez (2008, 153–155). Suárez does not directly reply to Pierson and Reiner, but to an equivalent argument by Hitchcock (1992, 160–166).

it blocks a possible reply to the conventionality objection, namely the causal realist's claim (in the spirit of Cartwright 1983, essay 4) that the difference between IMLC and IBTE is not just a matter of arbitrary convention, but firmly rooted in actual scientific practice. Not so, says Hitchcock, there are examples showing that scientific practice is actually opposed to treating IMLC differently from IBTE.

What are these examples? They both come from quantum mechanics, and since they are closely related to each other, it suffices to look at one of them. Consider the following well-known phenomenon: In a double-slit experiment with quantum objects (e. g., electrons), an interference pattern appears on the screen behind the two slits. But as soon as we perform a measurement to find out through which of the two slits each electron passed, the interference pattern disappears. This is in some textbooks explained by the following line of thought:

Suppose that we try to determine which slit a given electron passes through by bouncing a photon off of it as it passes through. The wavelength of the photon must be small enough that the position measurement on the electron after it has passed through the first screen is accurate relative to the distance between the two slits. The collision of the photon with the electron will impart some momentum to the electron in the vertical direction, with the amount of momentum transmitted being proportional to the frequency of the photon. If the electron receives a sufficient kick, its vertical displacement upon reaching the screen will be of the same order of magnitude as the distance between the crests of the interference pattern, so the interference pattern will be destroyed. (Hitchcock 1992, 170–171)

Hitchcock now claims that this explanation is *accepted* in the sense that it appears in many physics textbooks, but not *believed to be true*, because it contradicts some firmly established results from quantum mechanics, particularly the non-existence of simultaneously well-defined position and momentum values for electrons and photons and the resulting absence of classical particle trajectories. It thus seems that, as Psillos (2008, 169) puts it, Hitchcock raises “the spectre of van Fraassen”: While van Fraassen (1980) showed us that acceptance of *theoretical* explanations does not commit us to belief in their truth, Hitchcock now does the same for (at least some cases of) *causal* explanations.

But there is in fact no such alignment between van Fraassen and Hitchcock, and the illusion of their achieving the same thing turns on an equivocation of the term “accept”. Hitchcock uses the term in a very broad sense:

All of these causal stories are accepted in the sense that members of the scientific community deem them appropriate for the role they play; even the just-so stories can be accepted as serving important heuristic roles. (Hitchcock 1992, 174)

Van Fraassen's notion of acceptance is much narrower. According to him, *empirical adequacy* is a precondition for considering a theory acceptable (1980, 95). Now, is the

explanans in Hitchcock’s example empirically adequate? It surely “saves the phenomena” for this particular experiment, but we do not have to look far for phenomena which are incompatible with this classical picture; the interference pattern in the original double-slit experiment (*without* detection through which slit each electron passes) cannot be explained by classical electron trajectories and local interactions. Therefore, Hitchcock’s story can only serve as a counterexample to the validity of IMLC if we do not require causal explanations to be empirically adequate. But, as discussed in subsection 2.3, empirical adequacy is a necessary condition for causal warrant, so Hitchcock’s argument has no force against causal realism.¹⁷

4 Causal Reasoning in Context: The Case of the Neutrino¹⁸

4.1 Bohr and Pauli on Beta-decay

The problem that led Wolfgang Pauli to postulate the neutrino in 1930 was an observed inconsistency in the energy balance of nuclear beta-decay. The well-founded assumption that the nucleus has discrete energy levels before and after the decay conflicted squarely with the continuous energy spectrum of the emitted electrons. Pauli’s solution to this problem was that a hitherto unknown particle (the neutrino) might carry away some energy without itself being detected. Beside this hypothesis, there were other proposals to account for the continuous beta-spectrum, the most prominent being Niels Bohr’s idea that energy conservation might fail at the nuclear level.

Comparing these two proposals in the light of our discussion in section 2, the most obvious difference between them is that Pauli’s hypothesis meets criterion 2 (material inference) while Bohr’s explanation is clearly formal in character; it does not tell us what *causes* the continuity of the spectrum. This may be one of the reasons why Pauli, in a letter to Bohr in 1929, denied Bohr’s proposal even the status of an explanation: “So we really do *not* know what is going on here! You do not know it either, you can only give reasons *why* we do not understand anything.”¹⁹ On the other hand, its lack of specificity

¹⁷Suárez’s reply to Hitchcock’s objection is similar to mine, but since he does not formulate it in terms of empirical adequacy, it is not very convincing. Suárez denies (as I do) that Hitchcock’s example really is an acceptable causal explanation, but it seems that his only reason for doing so is that this causal story is not believed to be *true*:

The defender of causal explanation takes causal explanation to be a success term. So if we don’t believe in the ‘causes’ appealed to in the story then the explanation the story offers—if any—cannot be said to be causal. (Suárez 2008, 152)

But the point of Hitchcock’s argument was precisely to *deny* that causal explanation needs to be taken as a success term. So simply insisting on this linguistic practice will not do. In another passage, Suárez (2008, 159) seems to imply that the criterion of material inference helps to exclude Hitchcock’s counterexamples, but this will not work either, because these causal stories do not seem to depend on any formal inference. I conclude from this that the most effective reply to Hitchcock consists in requiring empirical adequacy as a criterion for causal warrant.

¹⁸For more details on this case study, see Egg (2012) and references therein.

¹⁹“Wir wissen also wirklich *nicht*, was da los ist! Du weißt es auch nicht, kannst nur Gründe dafür angeben, *warum* wir nichts verstehen” (Pauli 1979, 513; emphasis in the original).

saved Bohr’s hypothesis from any conflict with criterion 3 (empirical adequacy). The neutrino hypothesis, by contrast, had serious problems in this respect, as Pauli himself acknowledged in a letter to Oskar Klein in 1930: “So if the [neutrinos] really existed, it would hardly be understandable that they have not yet been observed. Therefore, I also do not myself entirely believe in the [neutrinos].”²⁰ This problem gradually disappeared in the years to come, when it became clear how weak the neutrino’s interaction with matter actually was.

With Bohr’s and Pauli’s hypotheses both (more or less) fulfilling criterion 3, there was obviously no causal warrant for the neutrino at that time, since criterion 1 (non-redundancy) was not satisfied. This assessment is not significantly altered even if one takes into account that the neutrino does not only ensure energy conservation in beta-decay, but also conservation of momentum and angular momentum; if one is willing to admit energy non-conservation, it is a small step to also allow for non-conservation of the other two quantities. However, it is an instructive thought experiment to ask with how much causal warrant we should credit the neutrino hypothesis on the basis of these observations, even if we assume (counterfactually) that criterion 1 had been satisfied by excluding Bohr’s hypothesis (and all other alternatives). Surely we would be warranted in inferring the reality of certain *properties*: the ability to carry away energy, momentum and angular momentum. But would we thereby be equally warranted to say that there is some *object* (the neutrino) which *has* these properties? The following analogy might illustrate this worry: Classical electrodynamics provided compelling evidence for the existence of electromagnetic waves, and it was only natural to assume that these were waves *of something*. But as it turned out, the underlying *something* (the ether) does not exist. May not the same apply to the neutrino? At this stage of enquiry, I see no reliable basis to refute this argument. One may claim that the simultaneous manifestation of three different properties (or at least two, given that the carrying away of energy and momentum might be viewed as the same property) provides some reason to believe in an underlying object having those properties, but this claim is not very convincing, since this simultaneous manifestation of properties has so far only been observed in the context of one kind of process, nuclear beta-decay. Hence the need (to be discussed in detail in the next two subsections) to detect the neutrino at a location other than that at which it was created.

4.2 The Impact of Fermi’s Theory and the Need for Direct Detection

In his *Versuch einer Theorie der β -Strahlen*, Enrico Fermi (1934) integrated Pauli’s rather qualitative neutrino hypothesis into a quantitative theoretical framework by describing the emission of an electron and a neutrino from a nucleus in close analogy with the emission of photons from an atom. The neutrino thereby became what is properly called a *theoretical entity*, and the impressive corroboration of Fermi’s theory in the years that

²⁰“Wenn die Neutronen also wirklich existieren würden, wäre es wohl kaum verständlich, daß man sie noch nicht beobachtet hat. Deshalb glaube ich auch selber nicht so ganz an die Neutronen” (Pauli 1985, 45). The neutrino was in that era still called “neutron”, since what we today call “neutron” was not discovered until 1932.

followed was by most physicists taken as confirming the existence of the neutrino. And even those who felt that a more direct proof was needed, had to recognize that “the theory was so attractive in its explanation of beta decay that belief in the neutrino as a ‘real’ entity was general” (Reines 1996, 317).

The question now is on what kind of warrant this general belief in the existence of the neutrino was based. Had the neutrino hypothesis somehow acquired causal warrant by being part of Fermi’s theory? I do not think so, for two reasons. First, the problem discussed at the end of the last subsection is obviously not affected by integrating the posited entity into a successful theory. Remember that the ether hypothesis was part of an immensely successful theory (classical electrodynamics), but it nevertheless turned out to be untenable.

Second, though it may seem as if the unmatched success of Fermi’s theory helped the neutrino hypothesis to fulfill criterion 1, a closer inspection shows that redundancy of explanations is still a problem. While it is true that, thanks to Fermi’s theory, the neutrino hypothesis proved to be incomparably more fruitful than Bohr’s hypothesis, this did not actually exclude energy non-conservation as a possible explanation of the continuous beta spectrum. Nor did it render Bohr’s hypothesis so implausible that adopting it would amount to general skepticism, as is the case with hypothesis H' discussed in subsection 2.1. Instead, the epistemic situation at that time is accurately described by the following quote from 1948:

While the [neutrino] hypothesis has had great usefulness, it should be kept in the back of one’s mind that it has not cleared up the basic mystery, and that such will continue to be the case until the neutrino is somehow caught at a distance from the emitting nucleus. Some physicists prefer to say simply that energy and momentum are apparently not conserved, giving full recognition, of course, to the energy and momentum relations that have been established experimentally, and to the success of the beta-ray theory which has been built upon the neutrino hypothesis. Perhaps all one can say is that this is a matter of taste. (Crane 1948, 278)

Another way of making this point is to say that the evidence that confirms the neutrino hypothesis via Fermi’s theory does so only indirectly (in the sense of Laudan and Leplin, as discussed in subsection 2.3), and as we have seen, this affords only theoretical, not causal warrant.

The dissatisfaction of some physicists with the credentials of the neutrino’s existence can thus be seen as resting on the distinction between theoretical and causal warrant. As long as a hypothetical posit has only theoretical warrant in its favor, skepticism about its actual existence is well founded, despite the success of the theory incorporating it. Accordingly, it is reasonable to stress the importance of the neutrino’s direct experimental detection. On the other hand, any philosophy failing to appreciate the distinction between theoretical and causal warrant will have to downplay this importance. The constructive empiricist, for example, will have to insist that a (so-called) neutrino detection can at most point to the *empirical adequacy* of the neutrino hypothesis, never to its *truth*.

But insofar as Fermi's theory already provides overwhelming evidence for the empirical adequacy of the neutrino hypothesis, the epistemic impact of a neutrino detection would be quite insignificant.

Similarly, an antirealist in the spirit of Stanford (2006) will deny that an experimental detection of the neutrino could justify belief in its existence, even if such an experiment managed to show that no other explanation (such as Bohr's hypothesis of energy non-conservation) accounted for the relevant phenomena. This is due to the so-called *problem of unconceived alternatives*: Elimination of all but one hypothesis is not sufficient to justify belief in this last contender, since there are always other possible hypotheses, which are not excluded, because no one has even thought of them yet.

It is not too surprising that antirealist positions tend to understate the importance of what physicists call a detection of some unobservable entity. But the same tendency is present in IBTE-based (or non-causal) versions of scientific realism, such as Psillos's. We have seen that, gauging by theoretical virtues, the neutrino hypothesis provided by far the best explanation of the observed phenomena. If no distinction between the strength of theoretical and causal warrant is drawn, then the overwhelming (theoretical) warrant in favor of the neutrino hypothesis already puts the existence of the neutrino beyond reasonable doubt, rendering its direct experimental detection superfluous.

Due to this consequence, non-causal realism is at odds with the convictions of a significant part of the scientific community, including the Nobel Committee, who awarded the 1995 Physics Prize to Frederick Reines "for the detection of the neutrino". This by itself is, of course, not a telling argument against non-causal realism; physicists may be mistaken about what is scientifically important and what is not. But if that was the case, then we would have to advocate major changes in the present and future priorities of particle physics, and this is in general not what scientific realists do.

The realist might try to bring his position into line with scientific practice by saying that the neutrino's detection is an important test, even though there already is high confidence in the correctness of the neutrino hypothesis, due to the success of Fermi's theory. Surprises, he might say, are always possible. But this amounts to conceding that the best explanation may not be good enough after all, and it invites the antirealist's question as to why this specific test (in contrast to all the previous tests) is supposed to take us from a state in which we are still prepared for surprises to a state of firm belief in the truth of the neutrino hypothesis. I will argue in the next subsection that the concept of causal warrant enables the causal realist to give a convincing answer to this question. But let me first dismiss a possible attempt to answer the question without relying on causal warrant. The non-causal realist might say that the significance of the neutrino's detection lies in the fact that it provides Fermi's theory with *novel predictive* success, while the theory could previously only boast *explanatory* success. This is a natural suggestion, given the centrality of novel predictions in contemporary arguments for scientific realism (see, e.g., Psillos 1999, 100–102). However, Fermi's theory had made successful novel predictions long before the neutrino was detected. For example, it predicted the possibility of *electron capture*, a process in which a proton in a nucleus absorbs one of the orbital electrons, forming a neutron and a neutrino. This process was subsequently detected by Luis Alvarez (1937). Therefore, we will have to look for

something else than just predictive success, if we want to grasp the significance of the neutrino detection experiments.

4.3 The Detection of the Neutrino by Reines and Cowan

In the mid-1950's Frederick Reines and Clyde Cowan performed a series of experiments by which they succeeded in directly detecting the neutrino. It may, of course, not be obvious what “directly” means in this context: neutrinos are neither visible themselves nor do they produce visible tracks in particle detectors. All that can be detected are the products of so-called *inverse beta-decay*, a process in which a proton absorbs a neutrino²¹ and turns into a neutron, emitting a positron: $\bar{\nu} + p \rightarrow \beta^+ + n$. These products initiate further processes, which then produce a characteristic double-peak signal in the particle detector. By 1956, Reines and Cowan concluded that at least some of these events were in fact produced by neutrinos.

So why was this a more direct proof of the neutrino's existence than earlier experiments? What exactly did Reines and Cowan achieve that was not achieved before? Given our previous discussion, it is natural to turn to criterion 1 and suspect that Reines and Cowan managed to eliminate redundancy by somehow excluding Bohr's hypothesis of energy non-conservation. And indeed, this hypothesis plays an important role in the assessment of their experiment, but it is a more delicate role than simply that of providing an alternative to the neutrino hypothesis. To see this, let us look at an extract from an article by Bruno Pontecorvo ([1946] 2000). He first summarizes the explanatory situation concerning beta-decay:

We see that the fundamental facts can be reconciled only with one of the following alternative assumptions:

- (1) The law of the conservation of the energy does not hold in a single β process.
- (2) The law of the conservation of the energy is valid, but a new hypothetical particle, undetectable in any calorimetric measurement—the neutrino—is emitted together with a β -particle in a β transition in such a way that the energy available in such a transition is shared between the electron and the neutrino. (Pontecorvo 2000, 23)

Then, after discussing a number of neutrino experiments, in particular some that included measurements of the recoil energy of the nucleus, Pontecorvo comments:

It should be noticed that experiments of this type, while of fundamental significance in the understanding of the β process, cannot bring decisive *direct* evidence on the basic assumption of the existence of the neutrino. This statement can be understood if we keep in mind that recoil experiments are interpreted on the basis of the laws of the conservation of the energy and

²¹More precisely, an antineutrino, but the difference between neutrino and antineutrino can be neglected in this context.

momentum in individual β processes, i.e., on the basis of alternative (2), which, in effect, corresponds essentially to the assumption of the existence of the neutrino.

Direct proof of the existence of the neutrino must, consequently, be based on experiments, the interpretation of which does not require the law of conservation of energy, i.e., on experiments in which some characteristic process produced by *free neutrinos* (a process produced by neutrinos after they have been emitted in a β disintegration) is observed. (Pontecorvo 2000, 24; emphasis in the original)

This is precisely what was done in Reines's and Cowan's experiment. What marks this experiment out as a direct detection is therefore not so much that it eliminates Bohr's hypothesis (1) as an alternative to the neutrino hypothesis (2), but that it provides evidence for (2) that does not itself presuppose the negation of (1).²²

But does the interpretation of the Reines-Cowan experiment really no longer depend on the validity of energy conservation? A diehard critic of (2) will deny this, claiming that the experiment can only imply the existence of the neutrino under the assumption that energy is conserved in *inverse* beta-decay. Without this assumption, we should expect inverse beta-decay to occur spontaneously, not needing a neutrino that produces it. Reines and Cowan would then simply have observed the process $p \rightarrow \beta^+ + n$, mistaking it for a neutrino-induced effect.

To counter this critique, a closer look at the experimental setup is necessary. Reines and Cowan did, of course, not just sit and wait for some inverse beta-decays to occur in their detector. Being aware of the exceedingly small cross section for this reaction (around 10^{-43} cm²), they set up their detector in a location where a high neutrino flux was to be expected, that is, next to a nuclear reactor. Now the interesting thing about a nuclear reactor is that it enables the experimenter to see if the event rate for inverse beta-decay changes when the reactor is turned on or off. And indeed, the signal rate was significantly higher when the reactor was operating than when it was shut down. It is here that the evidence for the neutrino becomes compelling, for the critic can only maintain his position by choosing one of the following options:

1. Deny that the correlation between the reactor status and the event rate suggests a causal link between the two.
2. Insist that the causal link between the reactor and the detector is mediated by something other than neutrinos. This may either be
 - a) some familiar objects, such as gamma rays or neutrons, or
 - b) some hitherto unknown entities.

²²This is a slightly different formulation of non-redundancy than the one given in subsection 2.1, and it has the advantage of describing more accurately what Reines and Cowan actually achieved. However, the two formulations are epistemically equivalent: independent evidence for the neutrino hypothesis amounts to evidence against the competing hypothesis.

Option 1 is a variant of the well-known phrase that “correlation does not imply causation”, but to invoke it here is to overlook the fact that Reines and Cowan did not just *observe a correlation* between reactor status and event rate, but rather found that turning on the reactor is an *effective strategy* to increase the rate of inverse beta-decay. If one were to deny that this establishes a causal link, one would also have to be skeptical about causation in everyday situations, for example that there is a causal link between regular toothbrushing and increased oral health. Such a level of skepticism would take us beyond the debate on scientific realism in much the same way as hypothesis H' discussed in subsection 2.1. Option 1 is therefore not to be taken seriously in the present context.

Option 2a, by contrast, has to be taken very seriously, and Reines and Cowan indeed took it so. They performed various tests to exclude any influence of known reactor products on the detector (see Reines et al. 1960). It is safe to say that they succeeded in excluding option 2a.

This leaves us with option 2b. Here we should first note that the critic can not just postulate a new entity (let us say, a *meutrino*) which has all the properties of the neutrino, but *is not* a neutrino. Given that there is nothing that makes a neutrino a neutrino apart from its properties, such a strategy would merely amount to *renaming* the neutrino. The critic thus has to make clear how the meutrino differs from the neutrino. But however he chooses to do that, the meutrino will have to share some properties with the neutrino, in order to explain the correlation between reactor status and event rate. In particular, the meutrino will have to be produced in the reactor and possess the ability to travel to the detector and produce an inverse beta-decay there. So even if we were to admit that Reines and Cowan might have detected meutrinos instead of neutrinos, it would still not have been the case that inverse beta decay occurred spontaneously, violating energy conservation.

In sum, we see that the evidence in favor of the neutrino hypothesis provided by the Reines-Cowan experiment does not depend on the validity of conservation laws at the nuclear level. Furthermore, it is clear that Bohr’s hypothesis now fails with respect to criterion 3, because it cannot explain the correlation between the status of the nuclear reactor and the rate of inverse beta decays in the detector. This implies that the neutrino hypothesis now fulfills criterion 1 and therefore, given that criterion 2 is still unproblematic, the hypothesis is now causally warranted.

Finally, notice that the neutrino’s properties no longer just manifest themselves in the context of its creation (in beta-decay), but also in the context of a second process that would not have occurred without the neutrino’s presence (inverse beta-decay). This dispels the worry (from subsection 4.1) that causal warrant may just extend to the *properties* of the neutrino, but not to the underlying *object* (the neutrino itself). It does not, of course, resolve the metaphysical question *what objects are*; they might be nothing more than bundles of properties. But some bond between the different neutrino-like properties is necessary in order to explain the causal link that Reines and Cowan found. *Something* must have travelled from the reactor to the detector, and this is what we call *neutrino*.

5 Conclusion

If scientific realism is to be plausible, it needs to take into account that not all scientific claims are equally well warranted; a differentiation of realist commitments is needed. I have argued that the distinction between theoretical and causal warrant provides a sustainable basis for such a differentiation. By introducing and justifying the criteria of non-redundancy, material inference and empirical adequacy, I showed that the distinction is not arbitrary and that causal warrant is indeed stronger than theoretical warrant. Furthermore, I defended causal realism, which is based on this distinction, against the objections which plagued a similar earlier version of scientific realism, namely entity realism. Finally, I illustrated the importance of causal warrant by studying the case of the neutrino hypothesis, where the neutrino's direct detection derives its significance from the fact that it endows the hypothesis, which was up to then only theoretically warranted, with causal warrant.

Of course, arguing for the strength of causal warrant *relative to theoretical warrant* does not go a long way towards defending causal realism against antirealism. I therefore conclude this paper by briefly assessing the prospects of causal realism to withstand the most telling antirealist arguments.

With respect to the *underdetermination of theories by evidence* (UTE), it was remarked in subsection 2.1 that many alleged cases of UTE do not have to be taken seriously, because they amount to arguments for radical skepticism, rather than (scientific) antirealism. The more general lesson to be drawn from this is that causal warrant must not be expected to yield infallible knowledge. At most, it can support the claim that our scientific knowledge is *no more* fallible than our everyday knowledge. This is the claim that the antirealist disputes by pointing to the various ways in which science involves a greater epistemic risk than common sense. For example, in a genuinely scientific case of UTE, it is risky to believe in any of the empirically equivalent theories. But these rare cases do not disturb the causal realist, because he simply admits that there is no causal warrant here, due to a failure of non-redundancy.

Another way in which science is riskier than common sense is that it makes claims about unobservables. In an attempt to avoid this risk, constructive empiricism suggests that science should not be seen as aiming for truth, but only for empirical adequacy. Van Fraassen (1980, 73) claims that this view “makes better sense of science, and of scientific activity, than realism does and does so without inflationary metaphysics”. There are some good arguments against this claim, which I will not rehearse here (see, e.g., Psillos 1999, chap. 9; Kitcher 2001, sec. 6). What I would like to repeat is an observation from subsection 4.2, namely that, in contrast to causal realism, neither constructive empiricism nor non-causal realism make good sense of scientific activity when it comes to the detection of the neutrino.

Finally, an important strand of antirealism rests on a *pessimistic induction* over the history of science, arguing from the falsehood of our best theories in the past to the probable falsehood of our best theories in the present. Currently, the strongest proponent of this type of reasoning is Stanford (2006), who uses it to argue for a specific kind of UTE, namely underdetermination between present and future (hitherto unconceived)

theories. Chakravartty (2008) sketches a reply to this argument by saying that those parts of our theories about which we have *detailed causal knowledge* are likely to be retained in future, so Stanford's UTE is no threat to them. My considerations about causal warrant are obviously related to this idea. In fact, I think that they are precisely what is needed to flesh out Chakravartty's sketch, but this will have to be shown in a different paper.

If these remarks are on the right track, then the significance of causal warrant for the scientific realist can be stated as follows: Claims that are causally warranted form the hard core of the realist's commitment, because they can be defended against even the strongest kinds of antirealist criticism (though not against radical skepticism, of course). This is not to advocate complete agnosticism with respect to claims that are only theoretically warranted. But we should acknowledge that such claims are significantly less secure than causally warranted claims, so they should be seen as forming a more peripheral part of the realist's commitment.

Acknowledgments Earlier versions of this paper were presented at research seminars at the universities of Zurich and Lausanne, at the 2010 Spring Meeting of the German Physical Society and at the workshop *Philosophy of Physics in Germany*, held in Hannover in June 2010. I am grateful to the audiences of these events for valuable discussions. I am also indebted to Michael Esfeld and two anonymous referees for many helpful comments.

References

- Alvarez, L. W. (1937). Nuclear K electron capture. *Physical Review* 52, 134–135.
- Bird, A. (2007). *Nature's Metaphysics: Laws and Properties*. Oxford: Oxford University Press.
- Cartwright, N. (1983). *How The Laws of Physics Lie*. Oxford: Clarendon Press.
- Chakravartty, A. (2007). *A Metaphysics for Scientific Realism: Knowing the Unobservable*. New York: Cambridge University Press.
- Chakravartty, A. (2008). What you don't know can't hurt you: Realism and the unconceived. *Philosophical Studies* 137, 149–158.
- Chang, H. (2003). Preservative realism and its discontents: Revisiting caloric. *Philosophy of Science* 70, 902–912.
- Crane, H. R. (1948). The energy and momentum relations in the beta-decay, and the search for the neutrino. *Reviews of Modern Physics* 20, 278–295.
- Devitt, M. (1997). *Realism and Truth* (2nd ed.). Princeton: Princeton University Press.
- Egg, M. (2012). Was bewirken Neutrinos? Eine Fallstudie zu Kausalität und Realismus in der Teilchenphysik. In M. Esfeld (Ed.), *Philosophie der Physik*, pp. 185–202. Frankfurt a. M.: Suhrkamp.

- Esfeld, M. (2010). Physics and causation. *Foundations of Physics* 40, 1597–1610.
- Fermi, E. (1934). Versuch einer Theorie der β -Strahlen. I. *Zeitschrift für Physik* 88, 161–177.
- Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.
- Hartmann, S., C. Hofer, and L. Bovens (Eds.) (2008). *Nancy Cartwright's Philosophy of Science*. New York: Routledge.
- Hitchcock, C. (1992). Causal explanation and scientific realism. *Erkenntnis* 37, 151–178.
- Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher and W. C. Salmon (Eds.), *Scientific Explanation*, Volume XIII of *Minnesota Studies in the Philosophy of Science*, pp. 410–505. Minneapolis: University of Minnesota Press.
- Kitcher, P. (2001). Real realism: The Galilean strategy. *The Philosophical Review* 110, 151–197.
- Kukla, A. (1993). Laudan, Leplin, empirical equivalence and underdetermination. *Analysis* 53, 1–7.
- Laudan, L. and J. Leplin (1991). Empirical equivalence and underdetermination. *The Journal of Philosophy* 88, 449–472.
- Morrison, M. (1994). Causes and contexts: The foundations of laser theory. *British Journal for the Philosophy of Science* 45, 127–151.
- Musgrave, A. (1996). Realism, truth and objectivity. In R. S. Cohen, R. Hilpinen, and Q. Renzong (Eds.), *Realism and Anti-Realism in the Philosophy of Science*, pp. 19–44. Dordrecht: Kluwer Academic Publishers.
- Norton, J. D. (2007). Causation as folk science. In H. Price and R. Corry (Eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, pp. 11–44. Oxford: Clarendon Press.
- Pauli, W. (1979). *Scientific correspondence with Bohr, Einstein, Heisenberg a. o.*, Volume 1: 1919-1929. New York: Springer.
- Pauli, W. (1985). *Scientific correspondence with Bohr, Einstein, Heisenberg a. o.*, Volume 2: 1930-1939. Berlin: Springer.
- Pierson, R. and R. Reiner (2008). Explanatory warrant for scientific realism. *Synthese* 161, 271–282.

- Pontecorvo, B. ([1946] 2000). Inverse β process. In K. Winter (Ed.), *Neutrino Physics* (2nd ed.), pp. 23–28. Cambridge: Cambridge University Press. First published as Report PD-205 (National Research Council of Canada, Division of Atomic Energy). Citations refer to the reprint.
- Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. London: Routledge.
- Psillos, S. (2008). Cartwright’s realist toil: From entities to capacities. See Hartmann et al. (2008), pp. 167–194.
- Reines, F. (1996). The neutrino: From poltergeist to particle. *Reviews of Modern Physics* 68, 317–327.
- Reines, F., C. L. Cowan, F. B. Harrison, A. D. McGuire, and H. W. Kruse (1960). Detection of the free antineutrino. *Physical Review* 117, 159–174.
- Stanford, P. K. (2001). Refusing the devil’s bargain: What kind of underdetermination should we take seriously? *Philosophy of Science* 68, S1–S12.
- 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. See Hartmann et al. (2008), pp. 137–163.
- van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: Clarendon Press.
- Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.