

How could models possibly provide how-possibly explanations?*

Philippe Verreault-Julien[†]

October 12, 2018

Abstract

One puzzle concerning highly idealized models is whether they explain. Some suggest they provide so-called ‘how-possibly explanations’. However, this raises an important question about the nature of how-possibly explanations, namely what distinguishes them from ‘normal’, or how-actually, explanations? I provide an account of how-possibly explanations that clarifies their nature in the context of solving the puzzle of model-based explanation. I argue that the modal notions of actuality and possibility provide the relevant dividing lines between how-possibly and how-actually explanations. Whereas how-possibly explanations establish claims of possible explanations, how-actually explanations establish claims of actual ones. Models, in turn, simply provide evidence for these claims.

1 Introduction

One puzzle, or paradox (Reiss, 2012), concerning highly idealized models is whether and under what conditions they explain. The paradox is that models

*I would like to thank audiences at the EIPE 20th Anniversary, INEM 2017, and EPSA 2017 conferences for helpful questions and comments. Special thanks also to James Grayot and the two anonymous referees of this journal for extensive and valuable suggestions on earlier versions of this article. The published version of this article will appear in *Studies in History and Philosophy of Science Part A*: <https://doi.org/10.1016/j.shpsa.2018.06.008>.

[†]Erasmus Institute for Philosophy and Economics, Erasmus University Rotterdam. Email: verreaultjulien@esphil.eur.nl. Website: <https://pvjulien.net/>. I acknowledge support from the Fonds de recherche Société et culture - Québec.

seem to misrepresent reality, and yet appear to be explanatory despite the fact that our best theories of scientific explanation (Woodward, 2003; Strevens, 2008) require faithful representation for successful explanation. For some commentators, one way out of this conundrum is to view these models as providing ‘how-possibly explanations’ (HPEs hereafter) (Craver, 2006; Forber, 2010; Reydon, 2012; Grüne-Yanoff, 2013a; Bokulich, 2014; Ylikoski and Awdmonat, 2014; Rice, 2016). They usually contrast HPEs to ‘how-actually explanations’ (HAEs hereafter), which are accounts of how phenomena actually occurred.

This response, however, raises two sets of issues. First, existing views attribute different features to HPEs. There are two important families of accounts, which I call the Dray-type and the Hempel-type. While the Dray-type (Dray, 1968; Forber, 2010) considers HPEs and HAEs to be a different species of explanation, the Hempel-type (Hempel, 1965; Brandon, 1990; Bokulich, 2014) distinguishes them by their degree of empirical support. These distinctions, I contend, do not appropriately capture the real contrast between them. To consider models simply as HPEs therefore does not straightforwardly solve the puzzle. Rather, it raises an important question about the nature of HPEs, namely, what demarcates them from HAEs?

The second issue concerns the relationship between highly idealized models and HPEs. Many models appear to not depict possibilities, but rather *impossibilities* (van Riel, 2015). If this is the case, then how could models possibly provide HPEs if their idealizations can’t possibly be true? Spelling out the epistemic contribution of models in terms of HPEs does not constitute genuine progress unless we have a clear idea of how models can indeed provide them.

I aim to provide an account of HPEs that clarifies their nature in the context of solving the puzzle of model-based explanation. More precisely, to address the first issue I argue that the modal notions of ‘actuality’ and ‘possibility’ provide the relevant dividing lines between HPEs and HAEs. The crucial feature that distinguishes HAEs from HPEs is neither the type of questions they answer nor the empirical truth of the former and the falsehood of the latter, but instead the sort of knowledge they provide. Whereas HAEs provide knowledge of *actual* explanations, HPEs provide knowledge of *possible* explanations. A HAE implies having knowledge about an actual state of affairs, whereas a HPE implies having knowledge of a possible one.

To address the second issue, I emphasize the need to make a distinction between model and world propositions. According to my account, HPEs are world propositions of the form ‘ $\diamond(p \text{ because } q)$ ’. What model propositions (e.g., unrealistic assumptions) do is to give reasons to believe in the truth of the possibility claim. In other words, they provide evidence for HPEs (cf. Claveau and Vergara Fernández, 2015). The prima facie issue of viewing models as HPEs dissolves when one properly takes into account the distinction between model and world propositions as well as the evidential relationship between models and explanations.

In section 2 I introduce the two types of HPEs. Section 3 shows why they are inadequate in the context of solving the puzzle of model-based explanation. I develop the account of HPEs as providing knowledge of possible explanations in section 4. In section 5 I spell out in more details what the relationship is between HPEs, models, and explanation. Section 6 concludes.

2 Two types of HPEs

The notion of ‘how-possibly explanation’ was introduced by Dray (1957) in his discussion over the adequacy of the deductive-nomological (D-N) model of explanation (Hempel and Oppenheim, 1948) in history.¹ Since Dray, the notion has been revisited many times and the concept was recently used in some discussions over the epistemic contribution of scientific models. Unfortunately, there is now a multitude of concepts on offer and it is now harder than ever to cash out a precise distinction between HAEs and HPEs. Table 1 gives a sense of the various terminology used when the terms are classified into three broad categories.

Table 1: Types of explanations

‘How actually’	‘How-possibly’	Misc.
why-necessarily (Dray, 1957)	how-possibly (Dray, 1957)	would-be explanation (Hindriks, 2013)

¹Dray (1954) did not yet use the terminology of ‘how-possibly’.

‘How actually’	‘How-possibly’	Misc.
true (Hempel, 1965)	potential (Hempel, 1965)	pseudo-explanation (Resnik, 1991)
how-actually (Dray, 1968)	possible explanations how-possibly (Dray, 1968)	how-plausibly (Machamer et al., 2000)
potential how-actually (Reydon, 2012)	global how-possibly (Forber, 2010)	genuine explanations in need of explananda (Reydon, 2012)
possible explanation why-necessarily (Dray, 1968)	local how-possibly (Forber, 2010)	more or less strongly confirmed (Hempel, 1965)

The literature thus far demonstrates that there is a relevant distinction between HPEs and HAEs; but where and how this distinction is conceived remains a source of contention (cf. Bokulich, 2014). How are we to make sense of this?

One possibility is to use the distinction between what Forber (2010) calls, following Sober (2003), Hempel’s problem and Peirce’s problem. The former is the problem of whether there are different types of explanations and of what constitutes an ideally complete explanation and the latter is the problem of comparing evidential support and accordingly choosing the best explanation. According to Forber, the dividing line between HAE and HPEs concerns Hempel’s problem, not Peirce’s. He uses the distinction to argue that since Brandon (1990) views the difference between HPEs and HAEs as one of evidential support—Peirce’s problem—, then what Brandon sees as HPEs are not, in fact, genuine ones. Rather, Forber says that Brandon’s HPEs should “count as incompletely confirmed or tentative how-actually explanations [...]” (2010, 36). Reydon agrees and call them “potential how-actually explanations” (2012, 307). These discussions indicate, as Reydon explicitly emphasizes, that HPEs and HAEs are sometimes believed to have a different logical structure. According to this view, HPEs and HAEs do not have the same form. A ‘potential how-actually explanation’ differs from a HPE in terms of form and also lacks the appropriate evidential support to be a genuine or complete HAE.

I find this discussion of Hempel’s and Peirce’s problems illuminating in that it suggests two dimensions along which to organize the accounts of HPEs on offer. As Strevens (2013, 512-513) notes, all accounts of explanation place two types of conditions on explanatory correctness: internal and external conditions.² Internal conditions state what *form* a set of propositions should take to count as an explanation. It is about the internal structure of the explanation. For instance, internal conditions for a D-N explanation indicate that the explanation needs to have the form of a deductive argument which takes a lawlike statement and initial conditions as its premises and an explanandum as its conclusion. The lawlike statement must be an essential part of the derivation. External conditions concern the empirical match between the explanation and what is to be explained. An externally correct D-N explanation, for example, is one whose lawlike statement, initial conditions, and explanandum are all true. It is important to note that whether these conditions are satisfied is in principle independent. An explanation can have the form of a D-N explanation, thus satisfying the internal conditions, yet be false. An explanation can also identify an externally correct explanans and explanandum, yet fail to connect them properly—e.g., the explanandum does not deductively follow from the explanans—and therefore does not satisfy the internal conditions.

Making this distinction allows us to see that there are in fact two main classifications in the literature. One class of HPEs states that the crucial difference between HPEs and HAEs is that they differ in their internal conditions. In the tradition of Dray (1957), these accounts suggest that HAEs and HPEs are, in virtue of a different logical structure, different species of the same genus explanation. Another class of HPEs instead states that the essential difference is in whether and how the external conditions are satisfied. Following Hempel (1965), these accounts state that HPEs are internally similar to HAEs, but they do not satisfy the external conditions for explanatoriness. To make this clearer, let us look at, respectively, what I call the Dray and Hempel types.

²De Regt (2009) makes a similar distinction between the “logical” and “empirical” requirements for scientific understanding as well as Hempel and Oppenheim (1948) with the logical and empirical conditions of adequacy.

2.1 The Dray-type: HPEs as a different species of explanation

Dray (1957) argued that there was a type of explanation that did not obey the strict logic of the D-N model. D-N explanations reveal why an explanandum *had to* happen, hence why Dray calls them ‘why-necessarily’. In explaining how-possibly, however, “we rebut the presumption that it [the explanandum] *could not* have happened, by showing that, in the light of certain further facts, there is after all no good reason for supposing that it could not have happened” (161, emphasis in original). According to Dray, the need for a HPE arises in the specific circumstances where one does not believe the explanandum could have happened. The HPE rebuts this belief by showing how it could possibly have happened. If someone believes it was impossible for something to happen, a HPE makes clear to her that it was, in fact, possible. It does not imply, however, that it had to happen. For Dray, a HPE reveals one or a few necessary conditions for the explanandum, though it does not show the sufficient conditions. However, this does not mean that HPEs are incomplete D-N explanations (cf. Brandon, 1990). Rather, they fulfil a different explanatory task, which is the refutation of the belief that some state of affairs is impossible. In this sense, HPEs can be complete. In D-N cases, if a belief is rebutted, it is the belief that the explanandum need not have happened. Put differently, what Dray calls why-necessarily explanations answer the question ‘Why it did so?’ whereas HPEs answer ‘How it could be?’ (Dray, 1968). For him, then, what distinguishes HPEs from HAEs is their different internal conditions: the “two kinds of explanation are logically independent” (Dray, 1957, 167).

In response to some critics, Dray (1968, 399) raises an interesting distinction between ‘possible explanations how-possibly’ and ‘possible explanations why-necessarily’. This makes clear that not all how-possibly explanations may count as true explanations. HPEs can also be false. Thus, a ‘possible explanation how-possibly’ is one that meets the internal conditions, but not the external ones. Likewise, a HAE can also only satisfy the internal conditions. In his terms, it is a ‘possible explanation why-necessarily’. ‘Possible explanations how-possibly’ and ‘possible explanations why-necessarily’ differ in terms of their internal conditions, but are similar insofar as they both do not meet the external conditions. For Dray, a HPE has to offer a true necessary condition. A HPE that would offer a

false or irrelevant one would only be a ‘possible explanation how-possibly’.

Forber (2010), following Dray, frames the problem similarly and identifies three different species of explanations: 1) global how-possibly, 2) local how-possibly, and 3) how-actually explanations. Like Dray, Forber argues that each sort of explanation answers a different question:

The first type [global HPE] answers the question: could some potential process produce evolutionary changes in idealized populations? The second [local HPE]: could some known process produce, in a way consistent with the local information set for a real population, an observed evolutionary outcome or pattern? And the third [HAE]: why, exactly, did some target evolutionary outcome or pattern occur (Forber, 2010, 34)?

For Forber, these explanations all have different internal conditions. Where Forber differs significantly from Dray, however, is when he claims that “we can successfully explain how-possibly without *any* empirical support [...]”; we need only show that some outcome or pattern is consistent with a specific set of information” (Forber, 2010, 39, emphasis in original). Indeed, Dray’s account requires that rebutting the impossibility presumption identifies an actual necessary condition for the explanandum. This of course demands some sort of empirical support. Relatedly, Reydon (2012), in a critique of Forber, also notes that his global and local HPEs do not exactly have the same structure as Dray’s HPEs. For Dray, HPEs serve to refute the belief that a certain state of affairs is impossible. This is not necessarily the case for Forber’s conception of HPEs. However, his project is similar to Dray’s in that he maintains that there are different *types* of explanations and that for him HPEs are not just incomplete or partial HAEs.³

2.2 The Hempel-type: HAEs lacking confirmation

Hempel denied HPEs were a different type of explanation. He held that ‘how-possibly’ concerned the pragmatics of explanation (Hempel, 1965, 428). The fact

³In a reply to Reydon (2012), Forber (2012) says this is the main feature which makes his project similar to Dray’s.

that an explanation may have to be presented differently depending on one's prior beliefs and explanatory requests does not imply that there are two different types of explanation. Accordingly, information may have to be presented differently in order to make it intelligible to different persons, but that does not imply that the explanations are fundamentally different.

Hempel, in contrast with Dray and Forber, rather argued that all explanations have to satisfy the same internal conditions, namely those specified by the D-N model. Yet, Hempel differentiated explanations according to whether they are 'potential', 'more or less strongly supported or confirmed by evidence', or 'true' explanations (Hempel, 1965, 338). A 'potential' explanation is defined as "any argument that has the character of a D-N explanation except that the sentences constituting its explanans need not be true" (ibid.). That is to say, it is an explanation that satisfies the internal conditions, but not the external ones—i.e., the explanans is false or not known to be true. A 'true' D-N explanation is one that meets both the internal and external conditions.

Hempel rejected the assertion that HPEs were a distinct species of explanation and therefore did not believe it was necessary to characterize them further. Even though some accounts (Salmon, 1989; Brandon, 1990; Craver, 2006; see also Bokulich, 2014) have equated Dray's HPEs with Hempel's potential explanations, it is in a sense misleading since for Hempel 'explaining how-possibly' belonged to the pragmatics of explanation and not to its logic. As it stands, Hempel's potential explanations can be viewed as another way of spelling out the distinction between HAEs and those 'explanations' that do not satisfy the conditions of HAEs. For the sake of the present discussion, what I call a Hempel-type HPE is thus a potential explanation.

Contrary to Dray, Hempel held that all explanations have the same internal conditions. Therefore, he rejected that there were different species of explanations. A potential explanation is not an explanation per se given that it does not actually explain. For a set of propositions to explain, the external conditions also need to be fulfilled. We can thus regard 'potential', 'more or less strongly supported by the evidence', and 'true' explanations as a continuum. Where they differ is in the degree to which they each satisfy the external conditions. The less evidence we have to warrant belief in the actual (true) explanation, the closer we are to a potential explanation. Conversely, if the explanation is for all practical

purposes confirmed, then we are closer to the actual (true) explanation. Were a Hempel-type HPE to be true—or be empirically confirmed—, then it would satisfy the external conditions and thus explain.

3 Dray and Hempel types in practice

Dray-type HPEs are thus distinct from Hempel-type potential explanations. A Dray-type HPE has both different internal and external conditions. Internally, it is a different species of explanation, whereas Hempel-type potential explanations are not. Externally, the strict Dray-type also needs to identify true necessary conditions for the explanandum. Potential explanations precisely do not have these truth requirements. It is thus rather surprising—and confusing—that some of the accounts in recent literature have adopted Dray’s language, but Hempel’s concept.⁴ For instance, Resnik says that “the difference between how-possibly and how-actually explanations is quantitative—a difference of degree—since empirical support comes in degrees” (1991, 143). To give another example, Craver maintains that how-possibly models “are purported to explain, but they are only loosely constrained conjectures about the mechanism that produces the explanandum phenomenon” (2006, 361). In these two cases, the use of HPEs is closer to the Hempel-type than to Dray. HPEs are not so much different types of explanation than ones not meeting the external conditions.

But is it not only a language problem? What is really at stake if scholars misidentify ‘potential’ explanations as ‘how-possibly’ explanations? The problem is that the confusion occurs not only at the semantic level. In many debates on theoretical models, the concept of HPE is mobilized to account for their epistemic contribution (Brandon, 1990; Cooper, 1996; Craver, 2006; Aydinonat, 2007; Grüne-Yanoff, 2013b,a; Rohwer and Rice, 2013; Bokulich, 2014; Ylikoski and Aydinonat, 2014; Rice, 2016). Unfortunately, the lack of a clear and shared account of HPEs makes resorting to them a shaky strategy. The following example illustrates the kind of issues we often encounter.

For instance, Rohwer and Rice (2013) argue that the Hawk-Dove model (May-

⁴Bokulich makes a similar point: “Subsequent discussions of how-possibly explanations typically treat them instead as *potential* explanations” (2014, 322, emphasis in original). See also Reydon (2012).

nard Smith and Price, 1973; Smith, 1982) is “explanatory” and provides “understanding”. It does so, according to them, even though it is *not* an explanation. It is explanatory and provides understanding, they say, because “the model is still able to answer certain key how-possibly questions (Resnik, 1991; Forber, 2010; Reydon, 2012)” (Rohwer and Rice, 2013, 349). In all fairness, Rohwer and Rice do not themselves develop extensively what they mean by ‘how-possibly’. Yet, the three authors Rohwer and Rice cite to support their claim all have substantially different views over what HPEs are. For Resnik, the main purpose of HPEs is heuristic. Fruitful HPEs can help to develop new and better theories. However, Resnik is silent with respect to the relation between HPEs and understanding. Forber, on the contrary, precisely argues against views such as Resnik’s by defending the view that HPEs, local and global, are a distinct type of explanation.⁵ This also contradicts Rohwer and Rice’s point that HPEs are not proper explanations. In fact, that HPEs *are* explanations in their own right is Forber’s main contention. Finally, Reydon argues, among other things, that Forber’s global HPEs are explanations, but not how-possibly. Without digressing into the details, as Rohwer and Rice describe the Hawk-Dove model it would clearly be a global HPE, not a local one. Thus, far from supporting Rohwer and Rice, this goes against their claim that the Hawk-Dove model is not an explanation and is only in the ‘how-possibly’ business.

As we can see, it is not possible to rely straightforwardly on the notion of HPEs to make sense of scientific modelling. Depending on one’s favourite account, HPEs can provide fruitful heuristics or be viewed as complete and genuine explanations to specific types of questions. This is of course making a significant difference when the question is precisely to assess the explanatoriness of models. If HPEs are to play a role in our appraisal of models, we better agree on what we take the features of HPEs to be and what their epistemic purpose is. Simply saying that models provide HPEs is unsatisfactory in the absence of an appropriate account.

Of course, that there are competing views over the appropriate notion of HPEs in the literature does not imply that all existing accounts of HPEs are mistaken. However, what I want to show is that neither the Dray-type nor the Hempel-type

⁵He considers Resnik to have a similar account as Brandon (1990), who he argues directly against.

are fully adequate. To do so, it suffices to briefly consider the case of Schelling's (1971; 1978) checkerboard model of residential segregation. The most popular interpretation of the model is that it provides a sort of HPE (Aydınonat, 2007; Grüne-Yanoff, 2013a; Weisberg, 2013; Ylikoski and Aydınonat, 2014). Weisberg (2013, 118-119) summarizes as follows the question the model raises and answers: "In other words, how is it possible for segregation to happen in a city without collective preferences for segregation? The answer is that this is possible when every individual has a small preference for similar neighbors and tries to satisfy this preference".

The model is not strictly speaking a Hempel-type potential explanation because it goes *beyond* meeting the internal conditions. The model does not only tell us how certain consequences can be derived. When considering the model, "we see Schelling's checkerboard cities as *possible cities*" and see the similarity between the model and the world "by accepting that the model world *could be* real - that it describes a state of affairs that is *credible* [...]" (Sugden, 2000, 25, emphasis in original). The model appears to tell us something more than just the fact that segregation can be obtained from given rules of behaviour. However, it is also not clear whether the model actually explains the general phenomenon of residential segregation or specific instances of it. For this reason, Aydınonat (2007, 430) considers whether we should view the checkerboard model as offering a "partial potential (theoretical) explanation". What the model shows is that it is possible that preferences for not living in a minority status bring about segregation. But that it is possible does not imply that it actually happens in this way. This is therefore different from offering a merely internally correct explanation and from providing a HAE. Furthermore, there is a *prima facie* puzzle on the standard reading of the Hempel-type since it was not developed to deal with the idealizations models typically contain. Laws of nature can be true or false, but we can't really say the same of models (Reiss, 2012). It is therefore unclear how the external conditions of the Hempel-type relate to models.

It is also not a Dray-type HPE for two different reasons. Firstly, what the model refutes is not the belief that residential segregation could not happen. The impossibility does not concern the explanandum, but rather the explanans. It was considered unlikely, if not impossible, that something other than strong discriminatory preferences could bring about residential segregation (Clark

and Fossett, 2008). As Grüne-Yanoff (2009) argues, the checkerboard model contradicts this belief. But no one had any reason to disbelieve that segregation exists or that it could happen. Secondly, since the checkerboard model does not identify actual necessary conditions for residential segregation, it would constitute a ‘possible explanation how-possibly’ (see Dray, 1968), not a HPE. Since the causal factors identified in the checkerboard model are at best sufficient for segregation—not necessary—and are not known to be actual, then it would only satisfy the internal conditions for a Dray-type HPE, but not the external ones. There might be other instances of segregation where these causal factors are not present.

4 The internal and external conditions of HPEs

What I take to be the distinguishing characteristic of HPEs is the modal information they convey. This appears to be a feature that both proponents of Dray-type and Hempel-type HPEs emphasize, albeit while not explicitly appealing to it. Consider the following quotations from commentators on diverse sides of the debate (emphases in original).

What we know seems to rule out the possibility of the occurrence which is to be explained. The explanation consists in showing that in spite of appearances to the contrary, it is not an impossible one after all (Dray, 1957, 161).

We use the notion of a potential explanation, for example, when we ask whether a novel and as yet untested law or theory would provide an explanation for some empirical phenomenon [...] (Hempel, 1965, 338).

What good is a speculative how-possibly explanation? The short answer is: it shows how known evolutionary mechanisms *could* produce known phenomena (Brandon, 1990, 180).

They [how-possibly models] describe how a set of parts and activities might be organized such that they exhibit the explanandum phenomenon (Craver, 2006, 361).

In contrast, how-possibly explanations aim to explain how some event could possibly occur (Forber, 2010, 33).

The Hawk–Dove game is intended to show how individual selection could *possibly* produce this behavior in a wide range of populations (Rice, 2016, 92).

All the preceding quotes suggest that HPEs have something to do with the *modality* of the explanation. The use of words like ‘might’, ‘could’, and ‘possibility’ are all modal terms. It is interesting to see that regardless of one’s specific position in the debate over HPEs, a common idea is that HPEs provide modal information. What I take to be the defining feature of HPEs is therefore not the type of question they answer nor their degree of empirical confirmation, but rather that they contribute to our knowledge of the possibility of certain states of affairs. HAEs, on the contrary, contribute to our knowledge of what is actually the case. Using this demarcation criterion, we can spell out in more detail a general characterization of the internal and external conditions of HPEs.

4.1 Internal conditions

The first question we may then ask is whether HPEs and HAEs have the same internal conditions, viz., do they have the same structure or form? In its general form, an explanation “is a set of propositions with a certain structure” (Strevens, 2013, 510). The question is thus whether HPEs and HAEs have the same propositions with the same structure. I propose that they do not. Following a useful characterization by van Riel (2015), I first take HAEs to express propositions of the following form:

HAE p because q

Here, ‘because’ is simply one’s favourite relation of explanatory entailment. We might say that ‘ p because q ’ on causal grounds, i.e., that ‘ q ’ is a cause of ‘ p ’. But ‘ p because q ’ could also be the case if a law of nature ‘ q ’ would be subsuming ‘ p ’. Nothing hinges on the specifics of what constitutes the relation. Whether explanations should have the form of a deductive argument or invariant counterfactual generalizations is not a relevant issue here. This general form is therefore flexible and can accommodate various substantial views over the

nature of explanation. A HAE is simply a set of propositions such that they are amenable to a formulation ‘ p because q ’.

One thing that is important to mention is that ‘ q ’ can be divided into several propositions, say the generalizations or laws of nature ‘ q ’ and the initial conditions ‘ c ’. A more precise characterization may thus be ‘ p because q and c ’. This will prove important when discussing specific accounts and can also make a difference when assessing, for instance, the epistemic contribution of a model. I will generally use the shorter version for the sake of brevity and simplicity. This characterization of HAEs should be fairly uncontroversial. The contentious issue is the following: How do HPEs differ from HAE with respect to form? Still following van Riel (2015), I suggest we view HPEs as propositions of the following form:

HPE $\diamond(p$ because $q)$

Propositions such as this mean “It is possible that ‘ p because q ’”. This captures a key desideratum of HPEs that virtually all accounts share, namely that HPEs express possibility claims. Propositions ‘ $\diamond(p$ because $q)$ ’ and ‘ p because q ’ do not have the same form and do not express the same content. Whereas knowing a HAE implies knowing what is *actual*, knowing a HPE entails knowing what is *possible*.

In this formulation, how the modality of the possibility operator should be interpreted is deliberately left open. This is an essential feature of the characterization I propose. Modality comes in various sorts, e.g., epistemic, metaphysical, causal, logical, nomic, etc. (see Kment, 2017). Quite often, the possibility of scientific interest is causal and typical HPEs thus establish that ‘ p because q ’ is causally possible. Logical possibility, on the other hand, is usually not very valuable insofar as logic rules out very few empirical possibilities. To know a logical possibility may not be informative with respect to learning about empirical the world. That said, as the burgeoning literature on mathematical explanation attests (Baron et al., 2017; Lange, 2013, 2017; Pincock, 2015), in some cases the relevant modality may well be mathematical. My goal here is not to restrict what counts as acceptable modality in all contexts, but rather to propose a general characterization. As I will show below, many accounts of HPEs rely on an implicit or explicit notion of modality, e.g., epistemic or causal possibility.

Leaving the interpretation of the possibility operator open has two main advantages. Firstly, it provides a flexible and unifying characterization. The same general schema can accommodate various HPEs. Causal or mathematical HPEs may have the same basic form, only different modal requirements. Secondly, it does not a priori assert what may be the relevant types of modality. In some contexts, scientists may be interested in nomic possibility. In others it may be mathematical possibility. My account does not preclude particular modalities.

How the possibility operator is interpreted has of course significant consequences on the external—or truth—conditions of a HPE. I will discuss it in more detail below, but here it is simply important to note that a thoroughly specified HPE states the modality of the operator. What matters is that the relevant modality ultimately needs to be selected in order to assess the possibility claim.

At first sight, it may appear that my proposal is simply a more formal characterization of Dray-type HPEs. Is this to say that, at least with respect to form, the Dray-type is the correct one? The Dray-type proponents are right to emphasize that the propositional content of HAEs and HPEs is different. But it is neither because HPEs answer a different type of question nor because they identify only necessary conditions. Firstly, for Dray (1968), HPEs only reveal necessary conditions and do not appeal to generalizations. Hence, the explanatory entailment relation—the ‘because’—is for him categorically different between HPEs and HAEs. In Dray’s (1968) parlance, if ‘ p because q ’ states sufficient conditions for ‘ p ’, then it is not a HPE, but rather an ‘explanation why-necessarily’. Dray only requires the identification of appropriate actual initial conditions ‘ c ’. A strict Dray-type could thus be formulated as follows: ‘ $\diamond p$ because c ’. The form is therefore quite different.

But to show that an explanandum is possible, it is not necessary to identify necessary conditions. Suitable sufficient conditions may do the trick. As Salmon (1989, 137) notes, “any potential explanation not ruled out by known facts is a suitable answer”. At the very least, some how-possibly questions can be successfully answered without citing necessary conditions. Moreover, as Reiner (1993) argues, finding necessary conditions is methodologically difficult, if not outright impossible. Indeed, the set of necessary conditions is considerably smaller since some causes may be individually sufficient for an explanandum. Establishing necessary conditions puts very high demands on our knowledge.

Perhaps there are HPEs that identify necessary conditions, but it seems we do not have good reasons to be monists about this. On the contrary, my account leaves open what is the set of possible interpretations of the explanatory ‘because’ relation.

Secondly, my account downplays the difference between HPEs and HAEs. One critical claim of the Dray-type is that HPEs and HAEs are different species of explanation. Both Dray (1957) and Forber (2010) construe HPEs as answers to different types of questions. For Dray, HPEs rebutted the belief in the impossibility of the explanandum. Similarly, as I showed in section 2, Forber frames his different type of HPEs as answers to different questions. However, as was previously objected (Pitt, 1959; Hempel, 1965), quite often one can answer a how-possibly question with a HAE. In fact, a HAE may be the best answer to a how-possibly question since one would not only know why something is possible, but why it is actually the case. Therefore, how-possible questions do not require a distinct type of explanation. A HPE can be sufficient for answering a how-possible question, but it is not necessary.

More concretely, why should we limit what questions can be asked in what contexts? Grüne-Yanoff (2013a) shows that in the course of constructing HPEs, scientists sometimes start with different questions or background information. One example he gives is Ainslie’s (2001) feedback model of self-control. The model indicates that preferences displaying hyperbolic discounting are compatible with the phenomenon of moderate impulsiveness through a process of recursive self-prediction. Here Ainslie shows from possible initial conditions how a possible process could generate an actual idealized phenomenon. While Grüne-Yanoff discusses Forber’s account, he does not describe this case as one of Forber’s local or global HPE. Indeed, this example does not seem to readily fit within the global vs local distinction. It is not an inquiry into the strictly formal constraints of psychological possibility (global) nor does it postulate a known process or concrete population. Yet, I think Grüne-Yanoff is right to view this case as a HPE. At any rate, the problem is not so much that the example could not fit within Forber’s categories, but that as useful as they are in describing particular cases of HPEs, they do not seem to naturally account for the variety of epistemic contexts.⁶ The characterization I propose provides a general characterization that

⁶Bokulich (2014) makes a similar claim using a case study on the ‘stripes’ of the tiger

situates differences in the varying interpretations of the possibility operator and in the actuality or possibility of the constituent propositions (e.g., explanandum or initial conditions) of HPEs.

In sum, I do not see any principled reason to restrict some questions and explanatory endeavours from qualifying as HPEs from the outset. Dray-type proponents have rightfully pointed out some specific instances of how-possible questions, but consideration of the diversity of practices and questions scientists address militate against setting strict constraints. As we have seen, sometimes what a model does is to establish the possibility of the process, not of the explanandum. More formally, one would have prior good reasons to believe that ‘ p ’, but not that ‘ q ’. A model may exhibit that it is in fact possible that ‘ $\diamond q$ ’ and therefore could increase confidence in ‘ $\diamond(p \text{ because } q)$ ’. We could also have cases where we have independent reasons to believe that p and q , but not that ‘ $\diamond(p \text{ because } q)$ ’. In this situation, a model may connect the dots between ‘ p ’ and ‘ q ’ and establish that ‘ $\diamond(p \text{ because } q)$ ’. The characterization of HPEs that I propose encompasses all forms of how-possible questions since it does not involve prior beliefs or explanatory aims. It also does not exclude that a HAE can provide a satisfactory answer to these question.

If there are substantial differences between the account I propose and the Dray-type, how is it different from the Hempel-type? One straightforward possibility is to simply reconstruct it as a false or unknown HAE: ‘ $p \text{ because } q$ ’, where ‘ q ’—and therefore ‘ $p \text{ because } q$ ’—is either false or not known to be true. For Hempel, what differentiated a potential explanation—the classic Hempel-type HPE—from a true explanation—a HAE—is that the latter fulfilled the external conditions whereas the former did not have to. Put differently, HPEs had the same form as HAEs, except for possibly being false or not known to be true. Understood as a potential explanation, a Hempel-type HPE thus does not tell whether or not ‘ $p \text{ because } q$ ’ is the case since it can be unknown. What it tells is how ‘ p ’, ‘ q ’, and ‘ $p \text{ because } q$ ’ stand in logical relation to one another. This characterization emphasizes that what separates HPEs from HAEs is whether the former satisfy the external conditions. Would they be met, i.e., were ‘ p ’ and ‘ q ’ to be true, then ‘ $p \text{ because } q$ ’ would be true. This also accords with more recent bush. She claims that global and local constraints can be mixed depending on the level of abstraction.

accounts of HPEs close to the Hempel-type. For instance, according to Brandon (1990) the main difference between HAEs and HPEs is one of evidential support, not of form. In other words, HPEs and HAEs have the same internal conditions. What demarcates them lies in the satisfaction of the external conditions.

The preceding analysis shows that there is a significant difference in terms of modal information between the account of HPEs I propose and the Hempel-type. The Hempel-type shows that a given set of propositions meets the internal conditions of adequacy, e.g., that the explanandum logically follows from the explanans. The main information it provides is one of logical explanatory entailment. This information, albeit necessary and valuable, is limited. Crucially, fulfilling the internal conditions of adequacy does not provide any information beyond the mere logical possibility of the explanation. It does not tell whether the explanation or its constituents—i.e., ‘*p*’ and ‘*q*’—are possible according to the typically interesting sorts of modalities like causal possibility. However, as testified by the quotes at the beginning of this section, modal information that goes beyond logical possibility appears essential to HPEs.

In fact, even accounts that have been traditionally classified in the Hempel-type put stronger modal demands on what should count as a HPE. For instance, both Salmon (1989, 137) and Brandon (1990, 178-179) consider consistency with known facts to be a hallmark of HPEs. On the Dray-type side, Forber (2010, 34) appeals to a “causal principle of possibility”. Furthermore, HPEs usually considered in the literature (Grüne-Yanoff, 2013a) do not simply show that there is logical entailment between the explanans and the explanandum. This is too easy. Instead, they also appear to provide information about the world.

Hence, perhaps a second way of characterizing the form of the Hempel-type HPE could be that it also has the form ‘ $\diamond(p \text{ because } q)$ ’, but the possibility operator should only receive a formal—e.g., logical—interpretation. This sort of possibility then leaves open whether the explanation is, for instance, causally possible. This characterization has the merit of highlighting the main issue with the Hempel-type HPE: it is too lax. As a matter of fact, a common worry in the HPEs literature (Brandon, 1990) concerns so-called ‘just-so’ stories, a reference to Rudyard Kipling’s (1912) fabulous origin stories (e.g., *How the Camel Got His Hump*) for children. We want to be able to demarcate between just-so stories that only have the form of explanations and genuine HPEs that

tell us something, perhaps limited, about the world. More generally, we want to be able to demarcate between true and false HPEs. There is a difference between ‘explaining’ the camel’s hump by citing its grumpiness expressed through repeated ‘Humph!’ and a HPE of the hump based on natural selection. While both may have the form of a HPE, only the evolutionary HPE is acceptable.

One challenge for accounts of HPEs is thus to allow for a demarcation criterion between genuine HPEs and ‘just-so’ stories. Just-so stories do not provide any substantial information about the world whereas HPEs inform us about the modal status of the explanation. This all suggests the need for stronger interpretations of the possibility operator. It may then be objected that my account does little in terms of a positive proposal to respond to worries raised by ‘just-so’ stories. While I indeed do not propose any substantive criterion, characterizing HPEs as propositions of the form ‘ $\diamond(p \text{ because } q)$ ’ is an important step. First, it allows us to accommodate various standards through the possibility operator. What sort of possibility scientists are interested in depends on the particular research setting. My goal here is not to a priori specify what possibilities are important across all contexts. Second, contrary to some strands of the Hempel-type, my account in principle allows for standards that go beyond meeting the internal conditions for explanation or beyond consistency with facts. Perhaps we need a more neutral label than ‘just-so stories’ for those sets of propositions that only satisfy the internal conditions of adequacy. My point is simply that a good account of HPEs should also have the resources to go beyond these. By allowing further modalities, especially causal, my account broadens the scope of the Hempel-type. Moreover, as we will see in the next section, it does so by making explicit that (true) HPEs also need to satisfy some external conditions of adequacy.

4.2 External conditions

Typically, the external conditions for HAE require that ‘ p ’, ‘ q ’, and ‘ p because q ’ are all true. More precisely, a realist account of explanation will require that ‘ p ’ and ‘ q ’ are true whereas an antirealist may be content if only ‘ p because q ’ is true under certain conditions. For an antirealist, the explanatory relationship may be true in the absence of true explanantia and explananda (see Khalifa,

2011).

Of course, if ‘ p because q ’ is true, then ‘ $\diamond(p$ because $q)$ ’ is trivially true. But the reverse is not true. A proposition ‘ $\diamond(p$ because $q)$ ’ may be true while the corresponding HAE ‘ p because q ’ is false. A HPE provides information about possibility whereas a HAE provides information about actuality. HPEs thus have more minimal empirical requirements. For ‘ $\diamond(p$ because $q)$ ’ to be true, none of ‘ p ’, ‘ q ’, or ‘ p because q ’ need to actually be the case.

HPEs are propositions that express that a given explanation is possible. HPEs may be true or false. Naturally, my goal is not to develop a full-fledged account of modality.⁷ However, let me state some relevant considerations. Here I assume, along with the others in the literature on HPEs that take modality as one key feature, that modal claims have a truth value and that it can be established. As I said earlier, what is possible or not depends on the interpretation of the possibility operator. What is logically possible may not be nomically possible and thus the ‘same’ proposition may be true under one modality and false under another.

For the realist, the position I will assume going forward, an explanation can only be possible if all of its constituents are also possible *under the same modal interpretation*. If either ‘ p ’ or ‘ q ’ are not only false, but also impossible, then it can’t be possible that ‘ p because q ’. For if a given explanans ‘ q ’ is causally impossible, then the corresponding causal HPE ‘ $\diamond(p$ because $q)$ ’ can’t be true. An impossible cause can’t possibly provide a causal explanation of a given explanandum. It could be logically possible, but not causally, hence the requisite to fix the interpretation of the possibility operator. Conversely, for ‘ $\diamond(p$ because $q)$ ’ to be true, both ‘ $\diamond p$ ’ and ‘ $\diamond q$ ’ need be true. Again, the possibility of a causal explanation requires the causal possibility of its constituent explanans. This also means that ‘ $\diamond(p$ because $q)$ ’ does not imply that p and q are not the case. The possibility operator may only specify the ‘because’ relation. In fact, it may be the case that in HPEs either p or q are true, or both. For instance, certain generalizations, initial conditions, and explanandum may be actual, but the possibility of the explanation may not have been assessed.

How does this view differ from existing accounts of HPEs? The account I propose

⁷For discussions on this topic, see, e.g., Gendler and Hawthorne (2002).

has different external conditions than the Hempel and Dray-types. The contrast is starkest with the Hempel-type where there are simply no requirements on the possibility of either ‘ p ’ or ‘ q ’. But by not requiring them to be possible, as I already indicated, a Hempel-type HPE can in principle contain all sorts of quirky explanantia and explananda. In this sense, the external conditions of the Hempel-type are looser than the ones I propose. In fact, it is a HPE precisely because it does not satisfy the external conditions. As long as it is not a HAE, but has the same form, we have a Hempel-type HPE. Crucially, as we have seen, a Hempel-type HPE may have the right form, yet this says nothing about the possibility of either ‘ q ’ or of ‘ p because q ’. The possibility of the statement ‘ p because q ’ is irrelevant. To know that ‘ $\diamond(p$ because $q)$ ’, however, requires that one makes a sort of modal assessment since ‘ p ’ or ‘ q ’ can’t simply be false, they must be possible. To make the judgment over the HPE, one thus has to bring in background knowledge.

Should we then just call HPEs, following Hempel, a ‘more or less strongly supported by the evidence’ explanation? In a sense, HPEs are more or less strongly supported by the evidence. But for Hempel, what is more or less supported by the evidence is the HAE, not the HPE. This is a crucial difference because the evidence we have may support judgments of possibility, but not of actuality. For instance, we may regard the checkerboard model as providing a HPE of residential segregation in the sense that it is, in fact, possible that the preferences for not living in a minority status cause segregation. Yet, we can also consider that we have no evidence that it is the HAE. In fact, we may even have evidence that some other cause is the HAE. In short, we may have zero evidence that the HPE is the actual explanation, yet still consider that it is possible. A HPE is therefore fully supported by the evidence for the specific claim it is making and this evidence may be independent of the evidence for the HAE.

The difference in external conditions with the Dray-type is more subtle as we can find various proposals for external conditions. However, some Dray-type HPEs put actuality constraints on either the explanandum or the explanans. On one side of the spectrum, as we have seen, for Dray (1968, 399-401) a HPE identifies an *actual* necessary condition for the explanandum. To show that a given phenomenon was not impossible, one could not only appeal to a possible

explanans. The necessary condition identified had to be actually instantiated. This necessary condition then rebutted the presumption of impossibility. According to Dray, if only a possible necessary condition was identified, then it had to be called a ‘possible explanation how-possibly’. In other words, it would have the form of a HPE, yet would not satisfy the external conditions. Under my account, a HPE may include actual elements, but does not require any.

On the other side of the spectrum, as stated earlier, Forber’s global HPEs, in particular, have minimal external conditions. He holds that mere consistency with formal constraints is sufficient and that no *additional* empirical support is necessary. On the contrary, his local HPEs have some actuality requirements. Only the initial conditions ‘*c*’ may be speculative, the explanandum and the generalizations being actual. He holds that local HPEs “*are just-so stories that speculate about the adaptive (or non-adaptive) evolutionary history of a lineage*” (Forber, 2010, 36, emphasis in original). For him, this is not per se a problem as long as they are recognized as such.

Contrary to some versions of the Hempel-type and Dray-type, the external conditions I propose are either more or less demanding. They can be more demanding insofar as the mere logical form of an explanation is often not sufficient. A stronger modal appraisal is called for. It requires one to evaluate whether it is true that the explanans and the explanandum are possible according to a relevant modality. Contrary to some versions of the Dray-type, they can be less demanding as my account does not require any actual proposition. HPEs can include claims of actuality, but actuality is not required to establish claims of possibility.

4.3 Applying the conditions: a summary

I would like to end this discussion of the internal and external conditions of HPEs with a brief synthesis of how it helps to organize our thinking about the existing literature. Its principal virtue is to show where the differences lie between versions of HPEs. One conclusion I reach is that contemporary accounts (Brandon, 1990; Forber, 2010) are closer in terms of form than they seem to be. Indeed, they all emphasize that the end product of a HPE is a possibility claim. In this respect, I would say that Brandon’s and Forber’s accounts are similar to

Account	Internal cond.	External conditions	
		Possibility	Actuality
Dray (1957)	$\diamond p$ because c	Epistemic (?) p	Necessary c
Hempel (1965)	p because q	Logical	—
Brandon (1990)	$\diamond(p$ because $q)$	Epistemic c , ‘because’	q
Forber (2010)	$\diamond(p$ because $q)$	Local causal c , ‘because’	p, q
Forber (2010)	$\diamond(p$ because $q)$	Global causal p, q, c	—

Table 2: An application of the conditions

mine in that they take HPEs to be propositions of the form ‘ $\diamond(p$ because $q)$ ’.

Reydon’s (2012) analysis of Forber’s (2010) account supports this claim. Indeed, Reydon argues that Forber’s global HPEs should be seen as “genuine explanations in need of explananda”, thus implying that they have the same form as HAEs. However, they do not meet the external conditions of HAEs.⁸ He also makes the case that Forber’s local HPEs are in fact similar to Brandon (1990) and Resnik (1991), that they all have the same “logical structure” (Reydon, 2012, 309). Finally, Reydon denies that local HPEs are a species of explanation—i.e., that they actually explain—because of a lack of (complete) empirical support.

We can find more variance in what external conditions the different accounts put forward. Brandon (1990) appears to say that HPEs should include actual generalizations ‘ q ’ and that they should be epistemically possible. Forber’s (2010) global HPEs should be causally possible—note the different modality—relative to the global information set, but imposes no actuality requirement. His local HPEs use the same modality, but relative to the local information set. And the explanandum and the generalizations should be actual.

Table 2 summarizes the preceding discussion, which suggests two things. First, that the internal conditions intersect in significant ways. The differences are less deep than we may think. Second, that most of the disagreements are located within the external conditions. What kind of modality matters and whether there are requirements about actuality are two important sources of contention.

⁸This view has similarities to Sugden’s (2011) ‘explanations in search of observations’, which I discuss in section 5.

5 HPEs, models, and explanation

Since many accounts of scientific modelling rely on HPEs to appraise models, we should have a good idea of whether they are a genuine species of explanation, what sort of epistemic benefits we may expect from them, and whether scientific models can indeed provide HPEs. To answer these questions, let us first look at the last issue, namely, the relationship between models and explanations.

Models *are not* explanations. Explanations are simply sets of propositions satisfying the internal and external conditions stated by one's favourite theory (Strevens, 2013). Whereas explanations are linguistic entities, models are widely viewed as being non-linguistic. This is clearly true of physical models such as the MONIAC (aka Phillips Hydraulic Computer), but also of the mathematical models usually discussed in the literature such as the Lotka-Volterra model (Weisberg, 2007). So, when we say that a model explains a given phenomenon, what this means is that there is a given model proposition—propositions that are true *within* the model—according to which '*p* because *q*' and that this proposition is epistemically related to the real-world proposition '*p* because *q*'. When we say 'Model *M* explains why *p* with *q*', we must be careful to not conclude that the model *is* the explanation. Instead, what it means is that the model provides sufficient reasons to believe that '*p* because *q*' is true. It is in this sense that models explain.⁹

Indeed, what many models do is to enable or justify the beliefs scientists have towards claims of this sort. Claveau and Vergara Fernández (2015) argue that models play an evidential role when model propositions make a difference to one's beliefs or justification for real-world propositions. When using a model, the model propositions one knows enter one's evidential network for real-world propositions. Sometimes, these model propositions make a difference by enabling belief in the real-world proposition or by increasing justification. For instance, they show, as illustrated by the Diamond-Mortensen-Pissarides (DMP) model of the labor market, that entertaining certain model propositions can for instance increase the justification one has in the real-world proposition that higher unemployment benefits lead to higher unemployment. But besides this specific example, the fact

⁹Explanations are always given on the background of prior beliefs and competing explanations and this does not imply a pragmatic or psychological account of explanation (see Woodward, 2003, sec 5.12).

that modelling is such a widespread—and successful—epistemic activity would be rather mysterious if model propositions did not enter and make a difference to scientists’ evidential networks. That models can and do provide evidence for propositions of the form ‘ $\diamond(p \text{ because } q)$ ’ or ‘ $p \text{ because } q$ ’ should hardly be controversial.

That said, one could ask in virtue of what *exactly* models can provide such evidence. For my current purposes, I do not want to take a particular stance on what constitutes these reasons to believe. It might be by virtue of the model standing in some suitable representation relationship with the world, e.g., of isomorphism (French and Ladyman, 1999) or similarity (Giere, 2010). The model might be a fiction that licenses inferences to the world (Suárez, 2009). Solving the (difficult) problem of how model propositions can correspond or license inferences to world propositions is naturally out of the scope of this paper. Nothing hinges on the specifics of how models provide these reasons; we only need to grant that models do play this evidential role.

Discussing one recent and influential account of models may help to see the evidential role of models and its relationship to possibility. Sugden (2000; 2009; 2011; 2013) argues, especially in his more recent work, that posited similarities between models and the world may license inductive inferences from the model to the world. For instance, Sugden describes the inductive schema of explanation where the facts that 1) the explanandum ‘ p ’ is caused by the explanans ‘ q ’ in the model world and that 2) both ‘ p ’ and ‘ q ’ occur in the world 3) provide “*reason to believe*” that ‘ p ’ is caused by ‘ q ’ (Sugden, 2000, 19, emphasis in original; 2013, 240). Here the model propositions clearly enter into the modeller’s evidential network. What happens in the model serves to justify the inference to the world. While similarity is, according to Sugden, the key notion to license inductive inferences, he argues that one important dimension along which to judge it is credibility.¹⁰ In this context, credibility means that the confidence we have in our inferences is “greater the extent to which we can understand the relevant model as a description of how the world *could be*” (Sugden, 2000, 24, emphasis in original). Credibility in that sense is not about considering that the model *is*

¹⁰Sugden recently made this point explicit: “The fundamental explanatory concept in my account of models is not credibility but similarity” (Sugden, 2013, 240). This puts into question to what extent his account is then different from ‘isolationists’ accounts like Mäki (2009), but this is out of the scope of the current paper.

real, but about judging that it is compatible with one's knowledge and beliefs about the world (Sugden, 2000, 25; 2009, 18; 2011, 718).

Using the account of HPEs developed here helps to clarify Sugden's views and the debate about them. In a discussion of Schelling's (1971; 1978) model of residential segregation, Sugden (2011) argues that the 1971 model is an "explanation in search of an observation" (2011, 722) whereas the 1978 model was really trying to tell us something about the world. In other words, the 1978 model is an explanation and the 1971 model is a potential explanation (Sugden, 2011, 734).¹¹ While Sugden does not really define what he means by 'potential explanation', for him the 1971 model is less credible than the 1978 one. It *could* be real, but not in the same manner. Or consider how Sugden argues against the 'conceptual exploration' (Hausman, 1992) view according to which theoretical modelling is about exploring the internal formal properties of models. Sugden believes that the checkerboard model, especially the 1978 one, and Akerlof's (1970) 'market for lemons' go beyond mere conceptual exploration. According to him, "they are sketches of processes which, according to their creators, might explain phenomena we can observe in the real world. *But the sense of 'might explain' here is not just the kind of logical possibility that could be discovered by conceptual exploration*" (Sugden, 2000, 11, my emphasis; 2009, 23).

My account provides a ready-made clarification of what distinguishes the 1971 and the 1978 models and of what demarcates conceptual exploration from other more ambitious modelling exercises. It locates the disagreement in the divergent interpretations of the possibility operator. In the terms I propose, Sugden argues that the checkerboard model goes beyond providing evidence for claims of logical possibility. While he does not frame the issue in terms of different modalities, it is clearly in the underlying argument. As explained earlier, his notion of credibility appears to be closer to a causal or epistemic interpretation of modality. Likewise, we could interpret the difference between the 1971 and 1978 models as one of modal appraisal. The 1971 model is not considered to be possible in the same sense as the 1978 one. Furthermore, this may also help us understand the source of the debate between Sugden and rival accounts (Aydinonat, 2007; Grüne-Yanoff, 2009; Mäki, 2009). Without clearly specifying

¹¹Sugden appears to use a pragmatic or instrumentalist account of explanation (see Sugden, 2013). Whether this is the right account and whether the checkerboard provides a HAE or not is disputed (cf. Aydinonat, 2007).

what sort of possibility commentators consider to be relevant for the case at hand, they run the risk of talking past each other. My account of HPEs is therefore not in disagreement with current views of models. It rather provides a framework within which we can think more clearly about the potential sources of disagreements.

One potential major objection to my account comes from van Riel (2015). He argues that models can't provide HPEs since many successful models do not describe possibilities, but rather *impossibilities*. His argument is that since many of the entities or processes postulated by models are conceptually impossible (e.g., models of water as particles or as continuous medium are inconsistent), then these models can't be considered to be HPEs. Indeed, many idealized entities and processes that models contain could not, *sensu stricto*, possibly be the case. Sugden, for instance, recognizes this state of affairs: "Economic models often contain idealisations which, if interpreted literally, *cannot* be true" (2009, 18, emphasis in original). But why is this not an obstacle?

The misunderstanding lies in taking model propositions at their face value. Of course, if the possibility operator is applied literally to them, then many, if not the vast majority, of models will depict impossibilities.¹² However, the objects of knowledge, HAEs and HPEs, are not propositions about the model, but about the world. A model does not provide a HPE in virtue of its model propositions being literally true of reality. Instead, it provides evidence for corresponding world propositions (Claveau and Vergara Fernández, 2015). Accordingly, we should not interpret the possibility operator as directly bearing on the model propositions. Possibility judgments are rather on the appropriate world propositions. For a given HPE of the form ' $\diamond(p$ because q)' to hold, it is thus not necessary that the model propositions of ' p ' or ' q ' are possible. It is simply required, as we have seen above, that these model propositions be part of one's evidential network for the HPE.

In fact, as Van Riel's (2015) own account of the content of model-based information suggests, it is unclear to what extent he would disagree with my proposal. Van Riel's project is to provide an account of the information supplied by models

¹²Van Riel (2015, 3845) is not concerned with logical possibility since it is not the possibility typically relevant for scientific usage. His argument concerns metaphysical, conceptual, epistemic, and nomological interpretations of modality.

that deviate—e.g., depict impossibilities—from reality. He suggests that using a hyper-intensional operator of the form ‘according to model M in context C , (p because q)’ allows propositions based on false models to be factive. To use his own example, we can’t *know* that the difference between day and night is due to the sun circulating around the earth since this is not actually the case. However, we can know that *according to the geocentric model*, this proposition is true. So it seems that Van Riel and I partly have different projects. He offers an account of model-based knowledge and holds that models *are* not HPEs, i.e., that models do not literally provide information about possibilities. On this we both agree. It seems the potential confusion stems from not properly distinguishing model propositions from real-world ones. But I see no reason in his account and elsewhere to deny that model-based propositions may provide evidence for real-world claims of the form ‘ $\diamond(p$ because q)’. If anything, his account lays the ground for a translation manual between model and real-world propositions—which I here assume the feasibility—and thus concurs with my views. I therefore do not believe that my account of HPEs is a target for his arguments. Some false models may provide reasons to believe that a given explanation is, or not, possible.

That said, there might be a connection between the degree of idealization of model propositions and their capacity to serve as evidence for world propositions. That models misrepresent and idealize is almost a platitude. Perhaps the ‘more impossible’ the model propositions, the less they can serve as evidence for world propositions. However, to what extent certain idealizations facilitate or hinder our capacity to learn about the world is an important, but separate, question. There is no a priori reason to doubt that some idealized models contribute to our knowledge. Again, we here only need to grant that some idealized models can indeed serve as evidence for world propositions. Bokulich (2014), for instance, argues that some highly idealized models may be closer to providing HAEs than others providing more concrete causal details. Nonetheless, it might also be the case that sometimes scientists do not adequately restrict their possibility judgments. They may incorrectly consider some world propositions to be possible. One benefit of the proposed account of HPEs is that it renders explicit the fact that possibility judgements are made. This can allow us to probe the interpretation of the modal operator—e.g., causal possibility or epistemic possibility—as well as the evidence provided for the judgements. If one considers

that a model provides evidence for a possibility claim, one needs to justify why one thinks so. Knowing what sort of claim is put forward is a first and significant step in its assessment.

Generalizing van Riel’s argument about impossibilities to cases of modelling impossible targets (Weisberg, 2013, sec. 7.2.2) may however raise another challenge for my account. For instance, Weisberg gives the example of a model of a perpetual motion machine. Perpetual motion machines are nomically impossible. For the machine to work, the physical laws would need to be different. These models can’t of course provide HAEs, but they also can’t provide HPEs, or so it seems, since the target is nomically impossible. Yet, it appears that modelling impossible targets is both a common practice and one that is epistemically valuable. Weisberg provides a general defence of modelling impossibilities:

Why should theorists who are primarily interested in studying what is actual try to understand what isn’t actual? The answer to this question cuts deep into the heart of theoretical practice: Theorists ultimately aim to partition the space of possibilities. They aim to understand what is possible, what is impossible, and why (Weisberg, 2013, 128-129).

According to Weisberg, modelling impossible targets allows us to learn about the contingency of actual states of affairs by investigating alternative possible histories. We can also learn their necessity by exploring what laws of nature would need to be different for a nomically impossible fact to be merely contingent. Suppose one thinks that a perpetual motion machine would be possible. Or suppose that one simply does not know why perpetual motion machines are impossible. By modelling one using actual physical laws, we can learn why, in fact, it is not possible. The laws nomically prevent the machine from working perpetually. We cannot have the laws of nature we have and a perpetual machine. Knowing impossibility claims of the form ‘ $\neg\Diamond(p \text{ because } q)$ ’ is equally valuable as knowing claims of possibility. If the proposition ‘ q ’ is nomically impossible, ‘ $\Diamond(p \text{ because } q)$ ’ can’t be true under a nomic interpretation. Some models thus give us true propositions of the form: ‘ $\neg\Diamond(p \text{ because } q)$ ’.¹³ Knowing possibilities is especially important on the background of beliefs of impossibility, whereas

¹³It should be noted that ‘ $\neg\Diamond(p \text{ because } q)$ ’ is logically equivalent to ‘ $\Box\neg(p \text{ because } q)$ ’, which means that ‘It is necessary that $\neg(p \text{ because } q)$ ’.

knowing impossibilities is significant on the background of beliefs of possibility (cf. Grüne-Yanoff, 2009). Both should be regarded as HPEs.

When we say that a model provides a HPE, what this means is that it provides reasons to believe that ‘ $\diamond(p \text{ because } q)$ ’. One important question we may then ask is whether HPEs thus defined constitute a genuine species of explanation, assuming that propositions such as ‘ $p \text{ because } q$ ’ are explanations. More generally, in what relation do HPEs stand with respect to HAEs? Put differently, HPEs do not explain actuality, but do they actually explain? I think one major reason why we want to settle this question is not so much for taxonomical reasons, but rather to know whether HPEs are epistemically similar to HAEs. This is especially important in the context of the puzzle of model-based explanation.

One reason we may want to reserve the genus ‘explanation’ for what we have so far called HAEs is because HPEs are unsatisfactory answers to explanation-seeking why-questions. If I ask why the financial crisis of 2007-8 occurred and one answers me in return that it was possibly caused by the greed of bankers, I will not be fully satisfied with the answer. I still will not know what was *actually* responsible for the crisis, which was the information my question asked for. Denying the explanation status to HPEs thus sends the clear signal that HAEs are what we ultimately care about. HPEs, all worthwhile as they are and perhaps the best we can achieve in certain circumstances, do not actually explain phenomena. Van Riel (2015, 3846) makes a similar point when he argues that since HPEs lack the actual ‘because’ explanatory entailment usually associated with explanations, then they should not be considered as such. Furthermore, it may also be philosophically simpler since we do not have to provide an account of what it is to ‘explain non-actuality’. Indeed, if HPEs are a genuine species of explanation, additional analysis would be needed to determine what is the explanans and explanandum of these explanations. However, even if we accept that HPEs are not HAEs—an uncontroversial point—, another crucial question that arises is whether HPEs make epistemic contributions comparable to HAEs. And if they do not, is it then appropriate to call these both ‘explanations’?

A reason we may want to say that HPEs are genuine explanations is because they may, like HAEs, afford understanding. An important epistemic goal of science is understanding. Science affords understanding of reality. One way it does so is by explaining phenomena. As van Riel notes, a prime candidate argument

for the view that HPEs afford understanding is to consider they are genuine explanations. For if they explain, then surely they also afford understanding. But it is also possible to make an argument to the effect that HPEs, while not actually explaining, afford understanding. Having a scientific explanation is a sufficient condition for scientific understanding. But is it necessary? And if propositions afford understanding, does this imply they should be qualified as an explanation? Put differently, that propositions explain is a cue that they afford understanding. However, if propositions afford understanding, is it equally a cue that they constitute an explanation? That HPEs may contribute to learning (Grüne-Yanoff, 2009, 2013a) or understanding (Rohwer and Rice, 2013; Ylikoski and Aydınoğlu, 2014; Rice, 2016) has some support in the literature.¹⁴ However, granting that HPEs afford understanding does not imply they should necessarily be viewed as a species of explanation. For one, Lipton (2009) argues that HPEs are a prime case of understanding *without* having an explanation. If Lipton is right, then HPEs may afford understanding without being explanations.

The previous analysis unfortunately does not provide a straight answer about whether or not HPEs should be viewed as a genuine species of explanation. What it does, however, is spell out the commitments that come with using the notion of HPEs. On the one hand, if HPEs are explanations, then it accounts for the fact that they explain or are explanatory. This would provide a straightforward answer to what sort of epistemic contribution models providing HPEs make. However, we would need to motivate further why they should share the genus ‘explanation’ with HAEs whose contribution is in many respects superior. On the other hand, if HPEs are not a species of explanation, this does not necessarily imply they can’t afford, for instance, learning or understanding. Nevertheless, more would need to be said concerning the value of knowledge of possibility. In particular, we need to understand better the conditions under which it can inform us about the world.

¹⁴Whether understanding can be obtained without a HAE is also subject to debate (Trout, 2002; Khalifa, 2013; Strevens, 2013).

6 Conclusion

Although prima facie obvious, that what demarcates HPEs from HAEs is the modal information they provide has been obscured in previous discussions over HPEs. This single criterion is both simple and clear while also acknowledging the idea that HPEs make a genuine epistemic contribution in the form of knowledge of possibility.

When all is said and done, both the Dray and Hempel-types proponents were on the right track concerning many features of HPEs. Supporters of the Dray-type were right to draw attention to the fact that HPEs may make a genuine epistemic contribution. The crucial difference between HAEs and HPEs is the end product, i.e., their particular propositional content, viz. claims of possible or actual explanations. Hempel-type advocates were right to downplay the difference between HPEs and HAEs insofar as HAEs also provide claims of possibility. In some cases, we may lack the empirical support to establish a claim of actuality, but may have enough for a possibility claim. In terms of form, they are not as radically different as the Dray-type would have liked us to believe.

This characterization of HPEs makes plain what it is for a model to provide a HPE: it provides evidence for propositions of the form ‘ $\diamond(p \text{ because } q)$ ’. Knowing the general form of HPEs and how models relate to them allows us to assess with more precision the contribution of particular models. Some models provide reasons to believe the explanans is possible, others the explanandum, or they may even provide evidence about their impossibility.

However, some questions remain open. Whether HPEs are a genuine species of explanation is one of them. At any rate, if we want the term ‘explanation’ to be reserved to HAEs—and there might be good reasons for this—, then we should better find a less confusing and more appropriate one for HPEs. While it appears safe to claim that HPEs may afford epistemic benefits in the form of learning or understanding, more work needs to be done in order to clarify how exactly knowledge of possibility can do it. But at least we now know better where to look and what to look for.

7 Acknowledgments

I would like to thank audiences at the EIPE 20th Anniversary, INEM 2017, and EPSA 2017 conferences for helpful questions and comments. Special thanks also to James Grayot and the two anonymous referees of this journal for extensive and valuable suggestions on earlier versions of this article.

References

- Ainslie, G. (2001). *Breakdown of Will*. Cambridge: Cambridge University Press.
- Akerlof, G. A. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics* 84(3), 488–500.
- Aydinonat, N. E. (2007). Models, conjectures and exploration: an analysis of Schelling’s checkerboard model of residential segregation. *Journal of Economic Methodology* 14(4), 429–454.
- Baron, S., M. Colyvan, and D. Ripley (2017). How mathematics can make a difference. *Philosophers’ Imprint* 17(3), 1–19.
- Bokulich, A. (2014). How the tiger bush got its stripes: ‘how possibly’ vs. ‘how actually’ model explanations. *The Monist* 97(3), 321–338.
- Brandon, R. N. (1990). *Adaptation and Environment*. Princeton: Princeton University Press.
- Clark, W. A. V. and M. Fossett (2008). Understanding the social context of the Schelling segregation model. *Proceedings of the National Academy of Sciences* 105(11), 4109–4114.
- Claveau, F. and M. Vergara Fernández (2015). Epistemic contributions of models: Conditions for propositional learning. *Perspectives on Science* 23(4), 405–423.
- Cooper, G. (1996). Theoretical Modeling and Biological Laws. *Philosophy of Science* 63, S28–S35.
- Craver, C. F. (2006). When mechanistic models explain. *Synthese* 153(3), 355–376.

- de Regt, H. W. (2009). The epistemic value of understanding. *Philosophy of Science* 76(5), 585–597.
- Dray, W. (1954). Explanatory narrative in history. *The Philosophical Quarterly* 4(14), 15–27.
- Dray, W. H. (1957). *Laws and explanation in history*. Oxford: Clarendon Press.
- Dray, W. H. (1968). On explaining how-possibly. *The Monist* 52(3), 390–407.
- Forber, P. (2010). Confirmation and explaining how possible. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 41(1), 32–40.
- Forber, P. (2012). Conjecture and explanation: A reply to Reydon. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43(1), 298–301.
- French, S. and J. Ladyman (1999). Reinflating the semantic approach. *International Studies in the Philosophy of Science* 13(2), 103–121.
- Gendler, T. S. and J. Hawthorne (Eds.) (2002). *Conceivability and Possibility*. Oxford: Oxford University Press.
- Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese* 172(2), 269–281.
- Grüne-Yanoff, T. (2009). Learning from minimal economic models. *Erkenntnis* 70(1), 81–99.
- Grüne-Yanoff, T. (2013a). Appraising models nonrepresentationally. *Philosophy of Science* 80(5), 850–861.
- Grüne-Yanoff, T. (2013b). Genuineness resolved: a reply to Reiss’ purported paradox. *Journal of Economic Methodology* 20(3), 255–261.
- Hausman, D. M. (1992). *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.
- Hempel, C. G. (1965). Aspects of scientific explanation. In *Aspects of Scientific Explanation: And Other Essays in the Philosophy of Science*, pp. 331–497. New York: Free Press.

- Hempel, C. G. and P. Oppenheim (1948). Studies in the logic of explanation. *Philosophy of Science* 15(2), 135–175.
- Hindriks, F. (2013). Explanation, understanding, and unrealistic models. *Studies in History and Philosophy of Science Part A* 44(3), 523–531.
- Khalifa, K. (2011). Understanding, knowledge, and scientific antirealism. *Grazer Philosophische Studien* 83(1), 93–112.
- Khalifa, K. (2013). The role of explanation in understanding. *The British Journal for the Philosophy of Science* 64(1), 161–187.
- Kipling, R. ([1902] 1912). *Just So Stories*. New York: Doubleday Page & Company.
- Kment, B. (2017). Varieties of Modality. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 ed.). Metaphysics Research Lab, Stanford University.
- Lange, M. (2013). What makes a scientific explanation distinctively mathematical? *The British Journal for the Philosophy of Science* 64(3), 485–511.
- Lange, M. (2017). *Because Without Cause: Non-Causal Explanations in Science and Mathematics*. Oxford University Press.
- Lipton, P. (2009). Understanding without explanation. In H. W. de Regt, S. Leonelli, and K. Eigner (Eds.), *Scientific Understanding. Philosophical Perspectives*, pp. 43–63. Pittsburgh: University of Pittsburgh Press.
- Machamer, P., L. Darden, and C. F. Craver (2000). Thinking about mechanisms. *Philosophy of Science* 67(1), 1–25.
- Mäki, U. (2009). MISSING the world. Models as isolations and credible surrogate systems. *Erkenntnis* 70(1), 29–43.
- Maynard Smith, J. and G. R. Price (1973). The Logic of Animal Conflict. *Nature* 246(5427), 15–18.
- Pincock, C. (2015). Abstract Explanations in Science. *The British Journal for the Philosophy of Science* 66(4), 857–882.

- Pitt, J. (1959). Generalizations in Historical Explanation. *The Journal of Philosophy* 56(13), 578–586.
- Reiner, R. (1993). Necessary conditions and explaining how-possibly. *The Philosophical Quarterly* 43(170), 58–69.
- Reiss, J. (2012). The explanation paradox. *Journal of Economic Methodology* 19(1), 43–62.
- Resnik, D. B. (1991). How-possibly explanations in biology. *Acta Biotheoretica* 39(2), 141–149.
- Reydon, T. A. (2012). How-possibly explanations as genuine explanations and helpful heuristics: A comment on Forber. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43(1), 302–310.
- Rice, C. C. (2016). Factive scientific understanding without accurate representation. *Biology & Philosophy* 31(1), 81–102.
- Rohwer, Y. and C. Rice (2013). Hypothetical pattern idealization and explanatory models. *Philosophy of Science* 80(3), 334–355.
- Salmon, W. C. (1989). *Four Decades of Scientific Explanation*. Pittsburgh: University of Pittsburgh Press.
- Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology* 1, 143–186.
- Schelling, T. C. (1978). *Micromotives and macrobehavior*. New York: W. W. Norton & Company.
- Smith, J. M. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
- Sober, E. (2003). Two Uses of Unification. In F. Stadler (Ed.), *The Vienna Circle and logical empiricism. Re-evaluation and future prospects*, Volume 10, pp. 205–216. New York: Kluwer Academic Publishers.
- Strevens, M. (2008). *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard University Press.

- Strevens, M. (2013). No understanding without explanation. *Studies in History and Philosophy of Science Part A* 44(3), 510–515.
- Suárez, M. (2009). Scientific Fictions as Rules of Inference. In M. Suárez (Ed.), *Fictions in Science: Philosophical Essays on Modeling and Idealization*, pp. 158–178. Routledge.
- Sugden, R. (2000). Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology* 7(1), 1–31.
- Sugden, R. (2009). Credible worlds, capacities and mechanisms. *Erkenntnis* 70(1), 3–27.
- Sugden, R. (2011). Explanations in search of observations. *Biology and Philosophy* 26(5), 717–736.
- Sugden, R. (2013). How fictional accounts can explain. *Journal of Economic Methodology* 20(3), 237–243.
- Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science* 69(2), 212–233.
- van Riel, R. (2015). The content of model-based information. *Synthese* 192(12), 3839–3858.
- Weisberg, M. (2007). Who is a modeler? *British Journal for the Philosophy of Science* 58, 207–233.
- Weisberg, M. (2013). *Simulation and similarity. Using models to understand the world*. Oxford: Oxford University Press.
- Woodward, J. (2003). *Making Things Happen. A Theory of Causal Explanation*. New York: Oxford University Press.
- Ylikoski, P. and N. E. Aydinonat (2014). Understanding with theoretical models. *Journal of Economic Methodology* 21(1), 19–36.