CAUSAL MODELS AND DEFAULT

Francesco Nappo

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Philosophy in the Department of Philosophy in the College of Arts and Sciences.

Chapel Hill
2016

Approved by:

Marc Lange

John Roberts

Matthew Kotzen

ABSTRACT

Francesco Nappo: Causal Models and Default
(Under the direction of Marc Lange)

This paper criticizes some recent arguments by Halpern (2008) and others to the effect that the structural equations framework used for modeling relations of singular causation ought to be supplemented with some function, a normality ordering, reflecting the distinction between default or normal states of things and deviant or abnormal ones. A central contention of this paper is that these proposals are insufficient to solve the problem they are intended to solve. This raises the question whether we should search for a replacement for the structural equations approach or instead think of the failure of the Halpern proposal as due to a deeper problem with the project. In the conclusion, I provide an argument in favor of the latter thesis.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

HP	Halpern and Pearl's (2005) account of singular causation
HPN	Halpern's (2008) account of singular causation
SEF	Structural Equations Framework

1. Introduction

After being successfully employed in many contexts in the natural and social sciences, the structural equations framework (SEF) has recently gained considerable attention in the philosophical literature on causation. SEF allows us to study the causal structure of a given situation by producing a model for that situation in which the various events are represented by appropriately chosen *variables* and in which the relations of dependence between the variables is governed by a set of *structural equations*. Seminal work by Pearl (2000) has suggested that the formal tools of SEF can be used to formulate a philosophical theory of *singular causation*, as exemplified by claims of the form 'event *c* was a cause of event *e*'. Put briefly, the idea is to understand the relations between variables in a SEF causal model as relations of counterfactual dependence and then to formalize a set of rules the application of which in any given situation should guarantee, when combined with an appropriate choice of variables, that the causal verdict determined by the rules (by which I just mean what the rules say about what was the cause of a particular event) accords with our causal judgment.

As we now know thanks to the work of Hiddleston (2005), Hall (2007) and Hitchcock (2007), accounts of singular causation based on structural equations must overcome two main problems. A first problem comes from the apparent arbitrariness in the choice of variables for a causal model. SEF accounts of singular causation provide no precise instruction about which aspects of a certain situation ought to be represented by a causal model in the form of variables. This seems problematic because, as several philosophers have pointed out (e.g. Halpern and Pearl 2005; Hall 2007), the verdicts of a causal model are fairly sensitive to differences in the number of variables adopted. Thus, whether SEF gives the right causal predictions appears to depend on an arbitrary choice concerning which and how many variables to include in the model. For convenience, I will refer to this problem for the SEF account as "the problem of the variables".

A second problem comes from the work of Hiddleston (2005) and Hall (2007). Again setting the details of their arguments aside for the moment, what Hiddleston and Hall have shown is that SEF accounts of singular causation suffer from an under-determination problem. More specifically, their examples demonstrate that there are pairs of circumstances that differ in their respective causal structures but such that their representations in SEF models are the same (or, to put it more accurately, they are *isomorphic*). As a result, SEF causal models do not seem able to differentiate between circumstances with distinct causal structures. The upshot seems to be that there must be more to causation than the structural equations and therefore that SEF accounts must be incorrect. For convenience, I will refer to this problem for the SEF account as "the problem of isomorphism".

In a series of papers starting with Halpern (2008), a new theory of singular causation based on structural equations has been advanced that purports to resolve the two problems above (cf. also Halpern and Hitchcock 2009; Halpern 2015). I will refer to this theory as HPN. Its proponents answer the problem of the variables by appealing to the psychological data indicating that singular causal claims are *subjective*: competent speakers can differ in their causal judgments even when they are presented with the same amount of evidence. According to HPN theorists, the subjectivity of causal judgments is nicely captured by what they call the *model-relativity* of the causal predictions in HPN. This notion corresponds to the idea that the causal verdicts provided by HPN are to be understood as relative to a particular choice of variables. HPN theorists claim that this model-relativity makes room for an elegant explanation of the disagreements between speakers about matters of causality, by partly tracking them back to a divergence in which aspects of the cases under consideration are being represented or emphasized. Let's call this the "model-relativity solution" to the problem of the variables.

To solve the problem of isomorphism, HPN theorists draw on data from psychology and experimental philosophy indicating that singular causal claims are sensitive to a distinction between "*default*" or "*normal*" states of affairs and "*deviant*" or "*abnormal*" ones. For instance, experiments by Knobe, Sinnott-Armstrong and others (2005) suggests that competent speakers are more likely to consider a person's action a cause of an event if their action breaks what is considered a moral or a social norm. Proponents of HPN use this and other examples to motivate the introduction in their

formal framework of a new function, a normality ordering, reflecting the distinction between default or normal states of affairs and deviant or abnormal ones. The introduction of this technical device avoids the problem of isomorphism, since the pairs of circumstances that have isomorphic causal models in the old SEF accounts have non-isomorphic models in the HPN account, due to differences in their respective default/normality orderings. For convenience, let's call this the "default-relativity solution" to the problem of isomorphism.

After briefly returning to the promised details, I will argue that HPN, which is arguably the most developed attempt that we have at an account of singular causation in the language of structural equations, is untenable. Importantly, the arguments that I will put forward do not turn on discussing the psychological data that proponents of HPN rely on, though it's worth signaling that it is a separate and open question whether their interpretation of the data is correct. Rather, I will focus on the more urgent philosophical problems faced by HPN as a theory of singular causation. On my interpretation, proponents of HPN have failed to draw the right conclusions from the problem of isomorphism. On their view, the problem of isomorphism constitutes evidence that notions such as default and normality ought to be part of our theory of singular causation. I will argue that the reality is quite different from that, since what the problem of isomorphism really shows is that neither HPN nor any relevantly similar theory couched in the language of structural equations can ultimately succeed in providing a complete account of singular causation. This does not imply that their causal models cannot constitute useful tools for representing causal structures; it only implies that it is a mistake to think we can turn the formal tools of representation provided by the structural equations framework into a full-fledged theory of singular causation.

Section five to eight contain the main argument for this conclusion. As the first part of the argument, in sections five and six I will bring out a set of technical difficulties facing the default-relativity solution to the problem of isomorphism. The main problems that I will raise against the introduction of a normality ranking are that it *i)* generates unclarities, *ii)* leads to predictions of conflicts of intuitions when our causal judgments are instead firm, and *iii)* it is subject to counterexamples (at least to the extent that this is possible for a theory that embraces the model-relativity of singular causal judgments). In sections seven and eight, I will then move the discussion to

a less technical and more philosophical level, with the purpose of explaining where these problems come from. By remarking on some of the examples raised earlier, I will then put forward my thesis that that neither HPN nor any relevantly similar theory couched in the language of structural equations can reasonably be considered a complete account of singular causal claims. In particular, I will argue that that reflection on the role of default and normality on our practices of causal judgment reveals that there are deeper facts about agreement and disagreement among competent speakers about matters of causality than what defenders of HPN are able to countenance.

The sections before the fifth contain introductory material. Section two summarizes the basic elements of the structural equations framework and its formal causal models. Section three introduces the problem of the variables and presents the model-relativity solution to it. Section four tackles the problem of isomorphism and the default-relativity solution devised by HPN theorists.

2. Causal Models

In the way of avoiding a laborious though mathematically more precise presentation of the SEF account of singular causation, I prefer to introduce the account by means of particular examples. Here is a relatively simple one:

> Overdetermination. Suzy and Billy each have a rock in their hands and decide to throw it at a window. The rocks arrive at the window at the exact same time, shattering it.

In this case, most competent speakers judge that both Suzy's throw and Billy's throw are causes of the window shattering. Hence we say that the shattering of the window is *over-determined* by Suzy's and Billy's actions. Respecting this judgment however turns out to be problematic for a family of views that analyze causation in terms of counterfactual dependence. On these views, *c* is a cause of *e* just in case, had *c* not occurred, *e* would not have occurred. This definition is problematic because it seems clear that Suzy and Billy caused the window to shatter and yet it is not true that, had Suzy not thrown her rock, the window would not have shattered, and also not true that, had Billy not thrown his

rock, the window would not have shattered (for more about counterfactual accounts of causation, cf. Collins, Hall and Paul 2004).

What makes SEF accounts of singular causation seem particularly attractive is that, unlike the simple counterfactual theories, they promise to deliver the correct causal judgment while preserving the intuitive association between causation and counterfactual dependence. The first step towards this result consists in building an apt causal model for the situation under study. Such model is obtained by specifying a set of variables, representing the events occurring in the situation, together with some structural equations representing the ways in which the events in question relate to or depend on each other. Formally, one begins by introducing the *signature* S for a causal model M, defined by a triple <U, V, R> where U represents the set of *exogenous* variables, whose values are determined by factors outside M, V the set of *endogenous* values, whose values are determined by factors inside M, and R is a function mapping every variable of M to a set of values (with at least two values for each variable). In the overdetermination case under study, natural choices for the endogenous variables are: ST (for "Suzy Throws"), BT (for "Billy Throws") and WS (for "Window Shatters). Hence, for example, the event of Suzy throwing the rock will be represented as: "ST=1"; the window shattering as: "WS=1".

One then defines the *causal model* M as the pair <S, F> where S is the signature and F is a set of *structural equations*, representing relations of counterfactual dependence between various possible values of the variables in the model. For instance, an apt causal model for the over-determination case will be given by the signature as defined above and by the following structural equations: [WS= 1 if either ST=1 or BT=1], roughly corresponding to the idea that the window would shatter if either Suzy or Billy were to throw. Since in our case the actual situation is such that ST=1 and BT=1, it follows that in the actual circumstances variable 'WS' also has value 1. However, it is important to note that the structural equations contain significantly more information than just a description of the actual scenario. In particular, the equations also contain information about what would happen if one of the variables in the models was assigned a value different from its actual one. For instance, from the equation [WS=1 if either ST=1 or BT=1] we can infer that, if as the result of a suitable intervention both ST and BT were set to value 0, then WS=0.

Once the variables and the correct structural equations are in place, we can see what, according to this model, is the cause of the window shattering. We do so by applying the following definition due to Halpern and Pearl (2005):

HP X=x *was a cause of* Y=y relative to model M iff:

AC1: The actual values of X and Y relative to M are x and y, respectively.

AC2: There is a partition of V (i.e. the set of endogenous variables) into two disjoint sets Z and W with X⊆Z and an assignment of values *x* and *w* to the variables in X and W, respectively, such that the following conditions hold:

(**AC2a**) If X is set to value x'≠x and W is set to value w then Y≠y.

(**AC2b**) If X is set to value x then Y=y even if W is set to value w and any subset Z'⊆Z is set to its actual value.

AC3: No subset of the variables in X satisfies AC1 and AC2.

Condition AC1 ensures that both the cause and the effect have in fact occurred. This is true in our case study: both Suzy's and Billy's rocks have been thrown and the window has shattered; hence ST=1 and BT=1 and WS=1. AC3 ensures inessential elements are not illegitimately introduced in X that have nothing to do with the fact that Y takes value y as the result of a causal process. This is trivial in our case because we are only working with two variables ST and BT, and so there are no subsets of the variables in X that we need to worry about.

Condition AC2a is a relative of the so-called "but-for" test for causation: *but for* the fact that the cause occurred, the effect would not have occurred. However, condition AC2a is importantly more permissive than the standard but-for test in that it allows for the dependence of the effect Y=y (in our case,WS=1) on the cause X=x (in our case, ST=1 and BT=1) to be tested under the contingency in which some other endogenous variable W in the same model is set to some value w possibly different from the actual one. This latter point is crucial for the correct assessment of the over-determination case. To see this, focus on Suzy's action for a moment. Although it is true that if Suzy hadn't thrown her rock, the window would have shattered anyway as the result of Billy's throw (that is, if ST=0 and

BT=1 then WS=1), it is also true that, if Suzy hadn't thrown her rock *and at the same time also Billy hadn't thrown his*, the window would not have shattered (that is, if ST=0 and BT=0 then WS=0). Thus condition AC2a allows for the dependence of the window shattering on Suzy's throw to come out once the variable for the event that Billy throws his rock is set to the non-actual value 0.

Condition AC2b then counteracts the permissiveness of AC2a by ensuring that setting W to the non-actual value w' has no effect on the dependence of Y=y on X=x. For instance, in the case under consideration AC2b ensures that setting the variable for Billy's throw to the non-actual value 0 has no effect whatsoever on the causal dependence of the window shattering on Suzy's throw, and hence that provisionally setting the variable for Billy's throw to 0 only serves the purpose of making sure that Billy's throw does not mask this otherwise evident causal dependence. Simple reflection on the over-determination case shows that both ST=1 and BT=1 satisfy AC2b. Since both ST=1 and BT=1 satisfy all four conditions above, they both count as causes of WS=1 according to the HP definition. Therefore, the HP definition correctly predicts the intuitive causal judgment that both Suzy and Billy were causes of the window shattering in the over-determination case.

A mathematically more precise description of the framework could be given but I prefer to skip it for favouring the discussion of what's philosophically more relevant. I will start by introducing the so-called "problem of the variables".

3. The Problem of the Variables

The problem can be initially characterized in the following way. In setting up the causal model for the overdetermination case, an appeal was made to the idea that the situation under study was "most naturally" represented by way of a simple model with 'ST' and 'BT' and 'WS' as endogenous variables. However, as many have pointed out (e.g. Halpern and Hitchcock 2010) no account of what counts as the correct or most "natural" choice of variables has yet been put forward in the literature on SEF. Moreover, it turns out that by adding or removing variables in the model (or by adding values for the variables to range over) one can alter the causal predictions. The easiest way to see this is to

imagine a model for the overdetermination case that simply does not contain a variable for Billy's throw. If the variable 'BT' is excluded, then obviously the resulting model (say M_1) would not predict that Billy's throw was a cause of the window shattering.

Previously, I have referred to this problem for SEF accounts as "the problem of the variables" and will continue using this terminology in the discussion to follow. The reactions to this problem have varied wildly. Some philosophers, such as Hall (2007), have argued that the problem of the arbitrariness in the choice of variables shows that the causal models employed in the structural equations framework are merely tools to represent an antecedently understood causal or counterfactual situation; on their view, it is a mistake to think that the structural equations could be used to produce anything like a full theory of singular causation. Defenders of the structural equations approach have argued for the opposite view. Far from seeing the problem of the variables as a fatal defect for a candidate philosophical theory of singular causation, they have claimed that it is instead a positive feature of the account. It is important to clarify their argument for this conclusion.

As a first step towards this argument, SEF theorists have pointed out that the absence of a precise criterion for an apt choice of variables would be a problem for their accounts only if their aim was that of exhaustively laying down the principles that competent speakers use to judge matters of causality among actual events. But SEF theorists reject the idea that this is what an account of singular causation should be aiming at. In support of this claim, they appeal to the evidence indicating that ordinary speakers can vary in their causal judgments even in the absence of any difference in information. For instance, Colingwood (1940) observes that after a car accident an engineer might cite as a cause of the accident the bad conditions of the road, while a police officer may bring up the speed of the cars involved. The two subjects might therefore agree on the physical details of the situation and still provide substantially different causal verdicts. Recent developments in psychology seem to corroborate this hypothesis by suggesting that people's moral beliefs can affect their causal judgments (cf. Cushman 2009; Cushman, Knobe, and Sinnott-Armstrong 2008; Hitchcock and Knobe 2009).

If singular causal claims are influenced by many context-dependent considerations, one might wonder what is the point of putting forward a theory of singular causation in the first place. The answer to this question constitutes the second part of the defense of SEF accounts. According to

various SEF theorists, the main point of a formal theory of singular causation is not to state the necessary and sufficient conditions for an event to be a cause of another event; rather the aim is to model *agreement* and *disagreement* among competent speakers about matters of causality. On their view, SEF accounts achieve this result by what they call the *model-relativity* of singular causation. This notion corresponds to the idea that the causal predictions in a SEF model are relative to a particular choice of variables. In their view, such model-relativity allows us to explain the variability in people's causal judgment at least in part in terms of divergences in which aspects of the causal structure under consideration are being represented or emphasized. For instance, if one subject judges that the event Y=y (say, the car accident) was caused by X=x (say, the road being in bad conditions), while another judges that it was caused by W=w (say, the speed of the car), then a SEF theory predicts that their disagreement might boil down to which aspects of the situation a person is representing or emphasizing in the two cases, corresponding to two alternative causal models of the same situation.

There seems to be an obvious problem with this, which is that the response ends up exaggerating the speaker's variability in matters of causality. After all, there are ever so many "unnatural" causal models for any single situation that we might want to represent, including the obviously inadequate model M_1 considered above. So it can be objected that one cannot rationally accept a framework that, as a matter of fact, appears to legitimize the use of such clearly inadequate models for the purposes of causal prediction. Against this worry, SEF theorists propose a tentative and deliberately vague list of rules that in their view can provide the basis for what they call a "rational critique" of unnatural causal models (cf. Halpern and Hitchcock 2005). The list includes recommendations such as:

(R1) the chosen variables must be sufficient in number to represent the essential features of the situation under consideration;

(R2) the variables to be included must represent only events that one is willing to take seriously as being part of the causal structure under study;

(R3) potential candidates for additional variables must not change the "topology" of the original model (cf. Halpern and Hitchcock 2009);

(R4) values allowed for each variable must not represent sets of events that bear logical relations to each other.

There are, to be fair, some unresolved questions in the vicinity concerning, for instance, what should count as a change in the "topology" of a causal model. These are difficult technical questions and the lack of an answer can be forgiven when a theory and a research project that are still in their early stages. Really, all that matters for our purposes is that the proposed rules can give us some guidance as to how to navigate the problem of the choice of variables in a causal model. For instance, R1 can be used to rule out M_1 as an inadequate model for the overdetermination case, since by lacking a variable corresponding to the event of Billy's throw, it misses out on an essential element of the situation under study.

4. Motivating Default

There is, however, a further problem for SEF theories of singular causation, the so-called problem of isomorphism. The challenge, initially raised by Hiddleston (2005) and Hall (2007) can be explained by considering cases of "bogus prevention". An example is:

> Assassin-Bodyguard. Assassin is about to put poison in Victim's drink. Bodyguard, anticipating Assassin, puts an antidote in Victim's drink. But Assassin has a last-minute change of mind and does not put any poison. Thus, Victim survives.

In this scenario, it seems clear that Bodyguard's action, even though it was intended to prevent a real threat from Assassin, was not in fact a cause of Victim's survival. This is why we call this sort of scenarios cases of "bogus prevention". The trouble for SEF accounts stems from the fact that an intuitively correct causal model for this scenario (say M_3) is isomorphic to M_0, the causal model for the over-determination case. To see the isomorphism, suppose that the variables in M_3 are: AA (standing for "Assassin does not put the poison"), BA (="Bodyguard puts the antidote") and VS (= "Victim survives"). Then little reflection shows that AA, BA and VS satisfy exactly the same

structural equations as ST, BT and WS in M_0. In particular, the following equation holds: [VS=1 if either AA=1 or BA=1], paralleling the equation [WS=1 if either ST=1 or BT=1] in M_0.

This isomorphism raises a problem because, while in the over-determination case the application of the HP definition to M_0 gives the correct result that Billy's throw is a cause of the window shattering, the same definition when applied to M_3 yields the intuitively incorrect verdict that Bodyguard actually saved Victim's life. Moreover, not only is it difficult to see M_3 as violating any of the conditions R1-R4 above, which means M_3 is unlikely to be dismissed as an illegitimate model, but the isomorphism of M_0 and M_3 is no isolated case. As Hall (2007) shows, there are several pairs of circumstances that intuitively differ in their respective causal structures but that the HP account is bound to treat on a par. For these reasons, the strategy of avoiding the isomorphism by insisting on alternative ways of modeling troublesome cases does not seem promising (but see section seven below). Indeed, even some of the most strenuous defenders of the SEF account concede that these cases of isomorphism show that "there must be more to causality than the structural equations" (Halpern and Hitchcock 2010, p. 18).

To cope with this problem, defenders of the SEF approach have argued that the causal models need to be made sensitive to a distinction between "default" or "normal" states and "deviant" or "abnormal" ones (cf. Halpern 2008, Halpern and Hitchcock 2010, Halpern 2015). As they point out, the idea that these notions have something to do with causation is not new to philosophers. Maudlin (2003), for instance, uses examples from physics to argue that our capacity for causal judgments is intimately connected with the ability to isolate what he calls "quasi-Newtonian" systems in which various assumptions are made about the default behavior of things and deviations thereof. Similarly, Menzies (2005) argues that our ideas about what's "normal" or "typical" influences ordinary judgments about omissions. For example, even though both the Gardener and the Queen do not water the plants, we tend to blame the Gardener, and not the Queen, for the omission since it's the job of the former to water the plants. Research by Knobe, Sinnott-Armstrong and others (2005) further builds on this case by showing that competent speakers are more likely to consider a person's action a cause of an event if this action breaks what is considered a moral or a social norm.

It is important to stress here that there are a variety of different notions involved in talk of "default" and "normality" in this literature. For instance, Maudlin's (2003) notion of default is concerned with the behavior that some things or systems display when not acted upon by external factors. As it relies on the concept of "acting upon", the notion of default in Maudlin's sense is an explicitly causal notion. On the other hand, the notion of "normality" invoked by Menzies concerns sometimes the idea of some event being statistically frequent (or "typical") and some other times the idea of some behavior being in accordance with the functional role assigned to a thing or system, or what the system in question is supposed to do or how it is supposed to behave. Still another notion of "normality" is the one that Knobe, Sinnott-Armstrong and others (2005) emphasize in their research, namely, the idea of some action being in accordance or in disaccordance with a social or moral norm.

On the view developed by Halpern (2008), Halpern and Hitchcock (2010), Halpern and Hitchcock (2015), our practices with respect to singular causal claims are best explained by assuming that competent speakers possess, in addition to a theory of causality, which is in their view given by the structural equations, a theory of "default" or "normality" encompassing all the notions mentioned above. Thus, according to this proposal, we should think of the problem of isomorphism as providing evidence, not that accounts of causation in the language of structural equations are incorrect, but rather that competent speakers rely on an implicit and presumably inchoate theory of what's normal or default in some given circumstance. As evidence for this claim, Halpern (2008) shows that by incorporating in the HP account the "missing" information about normality and default one can in effect provide a solution to the problem of isomorphism. The resulting account, which for convenience will be referred to as HPN, is one natural implementation of this idea in the formal framework.

The technical details of this solution consist in introducing a new function to the old HP account, called the "normality ranking" and defined in the following way. Let a world w be a complete description of the values of all the variables. We assume that each world is assigned a particular rank, which is some natural number. We stipulate that higher ranks correspond to less normal or typical worlds. Given a particular ranking function, that is, a function from worlds to natural numbers, a statement of the form "If p then typically q" is true if in all the worlds of least rank where p is true, q

is also true. Hence in a model where people do not typically put additional substances in somebody else's drink, the worlds where no additional substances are put have rank '0', those in which one substance is put have rank '1' and so on. An *extended causal model* is then defined as the triple M= [S, F, *k*], where [S,F] is a causal model and *k* is a ranking function. The definition of 'singular cause' is then exactly the same as in the HP account, with the only difference that in AC2(a) we require there be a world w_1 such that $k(w_1) \leq k(w_@)$ where $w_@$ is the actual world and the following holds at w_1: (X=x' and W=w). In other words, the only worlds we are allowed to consider in evaluating the counterfactuals in AC2(a) are those that are at least as normal as the actual world.

The proposal appears to deal nicely with the bogus prevention case above. To recall, M_3 is our causal model with random variables: AA (standing for "Assassin does not put the poison"), BA (="Bodyguard puts the antidote") and VS (= "Victim survives"), governed by the structural equation: [VS=1 if either AA=1 or BA=1]. We are now supposed to assign a rank to each statement of the form 'X=x' in our causal model. One natural way (though, mind you, not the only one) is as follows: worlds in which nothing is put in Victim's drink are the most normal; second come the worlds where either Bodyguard puts the antidote in Victim's drink or Assassin puts the poison (but not both); third comes the world where both Assassin and Bodyguard put something in Victim's drink. If we then ask if Bodyguard's action caused Victim to survive, we notice that BA=1 fails to meet the new condition AC2a. For now the only worlds we are allowed to consider are those that are at least as normal as the actual one, and in the counterfactual world w_1 which is just like the actual one except that in it BA=0 and AA=1, VS=1. Thus, Bodyguard does not count as a cause according to HPN.

The skeptical reader might wonder on the basis of which criteria do we decide which ranking is the most natural to adopt in a particular circumstance. Wrong question, it turns out. Compare: on the basis of which criteria do we decide which causal model is the most natural to adopt in a particular circumstance? HPN theorists give a similar answer to these two questions: just as when it comes to variables defenders of HPN appeal to the notion of *model-relativity*, so in this case they argue that singular causal claims are *default-relative*, that is, they are relative to a particular ranking of worlds on the basis of their normality or typicality. As a result, their view is that the causal verdicts resulting from a particular model must be understood as relative not only to a choice of variables but also to a

choice of normality ranking. However, in what follows, I will argue that the default-relativity solution to the problem of isomorphism is untenable.

5. Problems with Default-Relativity/1

Although it is difficult to deny that the cases presented by Halpern (2008), Halpern and Hitchcock (2009) make it plausible that the causal predictions of a model ought to be understood as relative to a default, there are a variety of examples in which it seems clear that the burden placed on our practices by the supposition of default-relativity is far too high. This raises the worry that default-relativity may not be the right kind of tool to respond to the problem of isomorphism. One source of concern is given by cases in which the default status of certain events is (to borrow the expression from Blanchard and Schaffer 2015) *underdetermined*. This phenomenon occurs whenever we lack any intuition or guidance with respect to the question of what the default status of an event is. An example of such cases is:

> Train. A train is traveling alongside a certain path. The engineers have built an automatic switch system such that, when the train arrives at a certain location, there is a 50% chance that it gets sent on the rail through the mountains and a 50% chance that it gets sent on the rail through the sea. This time, the rail through the sea is blocked, and so the passengers will arrive at their final destination only if the switch directs the train on the rail through the mountains. Fortunately for the passengers, the switch points to the rail through the mountains and so the train arrives at its final destination.

The intuitive causal judgment in this situation is supposed to be that the switch pointing to the mountain rail was a cause of the train's arrival at destination. This seems true regardless of the fact that we don't always consider automatic switch systems like the one described in this example as causes of the relevant effect. For instance, some philosophers, e.g. Hall (2007), think that in the case in which both rails are free from impediments, the fact that the switch pointed to the mountain rather than the sea rail was not a cause of the train's arrival at destination. However, the same philosophers

agree that, in cases where one of the rails is blocked, the switch pointing to the other rail was a cause of the train's arrival at final destination.

The problem brought up by this example is that there seems to be no clear way of providing a normality ranking for our extended causal model which we can justify independently of our intuitive causal judgment. For instance, in giving an extended causal model for the Train case one would need to assign a ranking to the events SM=1 and SM=0, the former standing for the event that the switch points to the mountain rail, the latter for the event that the switch points to the sea rail. The problem is that the fact that the switch points towards the mountains is neither a deviant nor a default behavior of the switch; nor is it statistically more frequent than its negation; nor is it in accordance (or, for that matter, in disaccordance) with a social or moral norm; nor is it where the switch is "supposed to" point to. Thus we don't seem to have any guidance as to how to go about selecting a particular normality ranking for our model.

One response to this problem would be that we should consider the normality rank of whole possible worlds instead of focusing on single events. But this suggestion, if anything, makes the problem worse. For suppose we adopt the variable 'TD' for designating the event of the train arriving at its final destination. Then it would seem that the most normal world is the one in which SM=1 and TD=1, since one might think it's unusual for trains not to arrive at their final destination. Hence one might argue that SM=1 and TD=1 must be more normal than SM=0 and TD=0. However, given condition Ac2a in the HPN definition of singular causation, which imposes that we only consider worlds that are at least as normal as the actual one when determining whether some event was a cause of another, it turns out that there are no more normal worlds where SM=0 and TD=0, and therefore SM=1 was not in fact a cause of TD=1 according to the new model. This seems intuitively wrong.

The only way to get the causal judgment right must therefore be to insist that the two worlds are on a par with respect to their normality ranking. Perhaps one might think this isn't too much of a problem, as we are dealing with cases of automatic switch systems where our intuitions may be fuzzy and unclear. It is important to note, however, that cases where the default status of certain events seems unclear prior to our causal judgment are pretty much ubiquitous. To give just another example, suppose that, as the train travels along the mountain rail, it meets a red light and has to stop. Suppose,

moreover, that if the train had traveled at a speed of 225 km/h instead of 250 km/h, it would not have encountered the red light and it would not have had to stop. It seems safe to say that traveling at 250 km/h instead of 225 km/h caused the train to find a red light and stop. But, once again, what should count as default and what should count as deviant in this scenario is unclear. The setup of the case does not seem to provide any basis for privileging 225 km/h over 250 km/h or vice versa. Presumably a defender of HPN would want to insist that 225 km/h is more normal than 250 km/h; or that the two possibilities are on a par with respect to normality/default. But, once again, it seems clear that this sort of move would enjoy no further motivation than the fact that we antecedently know our causal judgment to be a certain way.

A variation of the previous Train example gives us further insights into the problem that I have just raised:

> Engineers. As before, the engineers have built an automatic system that directs the train either on the mountain rail or on the sea rail at random. However, in this scenario the mountain rail is blocked, while the sea rail is not. The engineers can remove the block by activating a security system at the very same moment the train arrives at the location of the switch. When the train arrives at the location of the switch, it gets directed to the mountain rail. Fortunately for the passengers, the engineers have activated the security system at the right time, so the block is removed and the train arrives at its final destination.

Our intuitive judgment in this case is supposed to be that the activation of the security system was a cause of the train's arrival at its final destination. The question is: how do we build an extended causal model that captures our intuitive causal judgment? Let us set aside for a moment questions about reasonable choices of variables for this situation, and instead focus on the normality ranking that one ought to assign to the various events in question. We have already seen that there is some difficulty with assigning rankings to the event of the switch pointing one way or another. But when it comes to giving a ranking to the behavior of the engineers, matters become even more complicated: is it default for the engineers to remove blocks from train rails? Is it statistically normal for them to do so? Are they acting in accordance with a moral norm? Not only many of these questions are hard to answer,

but it seems that the most reasonable ways to answer them point towards opposite directions. For instance, removing blocks from train rails may be infrequent for an engineer but it may be among their normal functions ("what they are supposed to do"). Their behavior may be a deviance from a default (which is presumably to be at rest?) but it is in compliance with a legal norm.

As with the earlier Train case, one problem raised by Engineers is that the introduction of a normality ranking makes it difficult for us to select a causal model antecedently to knowing our causal judgment, because of the lack of intuitions and the conflicts between different notions of default and normality. This situation, it is important to stress, has no parallel in the case of choosing between alternative sets of variables for a causal model. For, unlike with the problem of the variables, in this case we seem to lack appropriate rules and principles that would guide us in the selection of a normality ranking for our causal model. Really, our only way to get the causal judgment right seems to be to start from the very causal judgment we are trying to predict and, so to speak, "reverse-engineer" an appropriate ranking. I believe this is the kind of problem for the HPN account that Halpern and Hitchcock (2009) are pointing to when they admit that "the introduction of normality exacerbates the problem of motivating and defending a particular choice of model", worry that is echoed in Blanchard and Schaffer (2015), who argue that HPN generates "complicating and unconstrained unclarities".

While I certainly agree with the concerns mentioned by these authors, I believe that there is a further puzzle raised by cases such as Engineers. In a nutshell, the puzzle is that, while the causal verdicts of the model are different depending on the normality ranking that we select, our causal judgments do not seem to be subject to a similar variability. For instance, suppose that the engineers do not trust the information they receive daily concerning whether the mountain rail is blocked, and for this reason they have a policy to activate the security system as a precautionary measure even when they are fairly sure that the mountain rail is not blocked. Arguably, making salient the fact that the engineer's activation of the security system is "typical" doesn't change our final causal judgment. Indeed, it seems that, even if to activate the security system is among the engineers' routine actions, it would still be the case that, on the day that the mountain rail is actually blocked, removing the block from the rail prevented a bad train accident. On the contrary, it is possible to show that, if we were to

build an extended causal model that attributes a lower rank to the event of the engineers activating the security system, reflecting the fact that the engineer's activation of the security system is "typical", the resulting causal verdict would be the exact opposite of our intuitive causal judgment, according to which the activation of the security system was in fact a cause of the train's arrival at final destination.

A defender of HPN may object at this point that, even in the modified scenario, there is a clear and perhaps overriding sense in which the activation of the security system constitutes a deviance from a default. Presumably, they would want to insist that our judgment reflects the idea that it is not "default" for engineers to intervene on the railway to ensure the safety of the passengers. Since this behavior constitutes a deviance from a default, they would argue that HPN rules that the activation of the security system was a cause of the continuation of the train's journey. However, these sorts of responses contribute little, if anything, to counter the point that I am making here. It is clear that, because of the variety of notions that factor in their normality ordering, HPN theorist can in principle take any of our intuitive causal judgments and, so to speak, "reverse-engineer" a normality ranking that suits their purposes and avoids the counterexamples. I concede that doing so is perfectly in line with their theory because, in light of default-relativity, one can always argue that the fact that our causal judgments go in one way or another is evidence that one or another notion of "default" is the most relevant in a particular circumstance.

However, escape-manoeuvres like the one just outlined are especially unsatisfactory in this particular context. In front of us is a theory, HPN, that makes certain kinds of predictions about under what circumstances competent speakers would disagree in their causal judgments, namely when their background views about what is "normal" or "default" don't match. If the theory was correct, one ought therefore to expect at least some noticeable variability in our causal judgments depending on the sorts of normality considerations that are most salient in those circumstances. Otherwise talk of "default-*relativity*" is just empty. Yet it looks as if no matter how hard we try to make certain normality consideration salient, by imagining situations in which it would seem natural to say that certain events or behaviors are "normal" or "default", we still seem unable to notice any change in our causal judgments concerning those situations. And although it is possible to insist that, in the imagined situations, there still remains an overriding sense in which the events or behaviors are

abnormal or deviant, thereby maintaining that there is some explanatory role for default-relativity to play even in those circumstances, that still doesn't address the fact that there seems to be a much simpler explanation of why our imaginative efforts don't bring about any variation in our causal judgments, which is that those causal judgments are simply not relative to a default.

At this point it should be clear that the whole defense of the default-relativity solution relies on the fact that in some other cases, e.g. Assassin-Bodyguard type cases, default information appears to play a substantive role in explaining our causal judgment, thus making default-relativity a somewhat valuable, even if burdensome, addition to our theory of singular causation. My aim in the following sections will be to argue that this way of defending HPN doesn't withstand scrutiny. In particular, in section six I will present a case of isomorphism that default-relativity doesn't seem to solve. Finally, in sections seven and eight I will argue that there is a much clearer and more straightforward explanation about why and how information about normality and default enters into our judgments in the cases emphasized by HPN theorists, which doesn't share the intuitive costs of the default-relativity solution and doesn't end up treating normality and default as anything like hidden parameters in our practices of causal judgment. Together, these arguments purport to show that default-relativity is the wrong way to understand what is going on in Assassin-Bodyguard and other isomorphism-generating cases.

6. Problems with Default-Relativity/2

To sum up what I have been arguing in the previous section, there seem to be at least two sorts of concerns raised by cases such as Train and Engineers. The first is that, because of the lack of intuitions and the conflicts between different notions of normality, we seem to lack any sort of rule or guidance to decide on the normality ranking to assign to the events in a given situation; we are instead forced to reverse-engineer a normality ranking for our causal model by inferring it from the fact that our causal judgment goes in one way or another. While I concede that this is compatible with the HPN theory, because it is possible that competent speakers may be oblivious to the relevant normality

considerations at play in some circumstances, the worry remains that the theory would prove totally unhelpful in applications, as there seem to be way too many cases in which our causal judgments seem stable and firm, and yet our best attempt at giving a causal model that matches our intuitive judgment cannot but to rely on the very judgment that we are trying to explain.

The second concern for the HPN theory is that it appears to make unsupported predictions about conflicts between causal judgments. If it were true that different normality rankings would ensue into different causal judgments, as default-relativity requires, then we would notice at least some variability in our judgments depending on the kind of normality considerations that are most salient in those circumstances. Yet this is not what we see in a number of cases; instead, what we see is that, in cases such as Engineers, our causal judgments are fairly stable regardless of the normality considerations that are most salient in those circumstances. What seems to be lacking, then, is a clear sense in which the salience of some specific normality consideration may be interpreted as in any way related to, not to mention responsible for, facts about convergence or divergence in our actual causal judgments. Indeed, without such a sense it is unclear how to even assess the claim that singular causal claims are relative to a default.

In light of the previous discussion, it seems unclear to what extent one could argue that what is missing from the original HP account of singular causation is specifically and solely information about default, normality, typicality etc., since as we have just seen this kind of information appears relevant only in some cases, while it seems to be completely irrelevant in others. This is unfortunate because the interest in the HPN proposal was motivated by the promise of unifying the independently grounded psychological evidence with the philosophical theory, and the examples above seem to show that even if there is an influence by considerations of default and normality, this influence is only apparent in a limited range of cases (e.g. the case of the Gardener and the Queen, or the Assassin and the Bodyguard). In what follows, my aim is to argue against another crucial claim that HPN theorists have defended, namely that introducing a normality ranking into the original HP account solves the problem of isomorphism. Following Gallow (2015), I will use cases of so-called "preventative preemption" to argue that default-relativity is insufficient to solve the problem of isomorphism.

‘Preventative preemption’ is a technical term for a not-too-unfamilar kind of situation. An example of such situation was discussed in a seminal paper by McDermott (1995), who introduces the case in this way:

> Suppose I reach out and catch a passing cricket ball. The next thing along in the ball’s direction of motion was a solid brick wall. Beyond that was a window. Did my action prevent the ball hitting the window? (p. 524)

There seem to be two contrasting intuitions. Perhaps the most immediate answer that comes to mind is “yes, my action prevented the ball hitting the window”. However, on reflection it seems also clear that the window would never have been hit because of the presence of the solid brick wall. So perhaps we should consider saying: “no, my action did not prevent the ball from hitting the window”. Notice: this is a case where which causal judgment is correct may be intuitively unclear (even though I must register a pretty strong propensity to answer ‘yes’ on reflection). In any case, there seem to be certain aspects of this scenario that are intuitively less difficult to assess. In particular, we wouldn’t want to say that the solid brick wall prevented the ball from hitting the window, since the ball was intercepted by my catch before even hitting the wall.

Gallow (2015) has noted that cases of preventative preemption present a problem for the default-relativity solution to the problem of isomorphism. While I think that the specific example he uses is ultimately unsuccessful, I believe that his basic idea is correct. My preferred example is a variation on McDermott’s. Suppose that Suzy and Billy are at some distance apart and, like every Sunday, they are passing a ball to each other. Unbeknownst to them, a team of engineers have built a transparent partition in between the two that can be raised up at will so as to prevent the ball to reach the other player. Suppose that Suzy throws the ball with the intention of reaching Billy, who is ready for the catch. Suppose, further, that the partition is currently up, so there is no way Billy will receive the ball (even though he would have received the ball if the partition was down). The ball is about to be stopped by the transparent partition when Billy’s dog intervenes with a fantastic jump and catches the ball before it hits the partition.

Did Billy's dog prevented the ball from reaching Billy on the other side of the field? As in McDermott's example, it may be unclear what answer we should give to this question (though, once again, my intuition says 'yes'). In any case, we seem to be somewhat more certain of the fact that the transparent partition was up when Suzy threw did not prevent the ball from reaching Billy ("all that money and time spent for nothing!"). The problem for HPN is that a natural causal model for this situation ends up predicting that the transparent partition was instead responsible for the fact that Billy did not receive the ball. The model I have in mind is given by the variables 'DC', standing for the event that Billy's dog reaches out and catches the ball, 'PU', standing for the event of the partition being up, and 'BC' standing for the event of Billy receiving the cricket ball. The structural equations are then written down as follows: [BC=1 iff PU=0 and DC=0], meaning that Billy would have received the ball if and only if the partition was down and Billy's dog didn't make such a fantastic jump and catched the ball.

The crucial issue turns on assigning a normality ranking to the events involved. Notice that in this case we don't have a determinate causal judgment to rely on, from which we can reverse-engineer the normality ordering. Instead, we seem to have pretty strong intuitions about which events in the causal story under considerations can be considered default or normal, and which can be considered deviant. In particular, there seems to be no doubt that the presence of a transparent partition in the middle of a field on a Sunday morning would be a pretty exceptional event. Similarly, the dog's jump and catch is also quite uncommon. Hence there seems to be little doubt that the actual world, where DC=1, PU=1 and BC=0, is very abnormal. Slightly less abnormal are the worlds where either the dog catches or the ball hits the partition; the most normal world is the one where neither of these things happen and Billy receives the ball as usual.

Did the partition being up prevent the ball from reaching Billy? The HPN definition says that it did. It is not difficult to see why. First of all, PU=1 and BC=0 are all actual events, so AC1 is satisfied. AC3 is also satisfied because there are no intermediate variables to consider. Moreover, it is possible to partition the variables in two sets, with PU on one side and the rest on the other; and it is true that there is a more normal world than the actual one where PU=0 and BC=1, namely the world where DC=0. Such a world is a witness for the counterfactual: if PU=0 and DC=0 then BC=1.

Therefore AC2a holds. Similarly, it is true that if PU=1 then BC=0 even though DC=0. Therefore AC2b is satisfied. As a result, the HPN definition rules the transparent partition to have prevented the ball from reaching Billy. And given that HPN also rules the dog's catch as preventing the ball from reaching Billy, it follows that the HPN account treats this case of preventative preemption as a case of over-determination. Indeed, if we substitute the variable 'BC' in the model for the variable 'BD', standing for 'Billy doesn't catch', it is fairly easy to see that the resulting model is isomorphic to the model M_0 (the model for the overdetermination case described in section two).

Consequently, what the example seems to show is that the appeal to default and normality doesn't solve all the problematic cases of isomorphism. I can imagine two kinds of responses to this argument. A first response would be to insist on a different normality ranking for the events involved. Even though this is a genuine possibility, it is certainly not attractive. One interesting feature of the example just given is that, while there may be some uncertainty about the final causal judgment, we seem to have pretty clear and strong intuitions about what would count as default and what would count as deviant in those circumstances. To insist on a different ranking would therefore amount to an *ad hoc* rejection of some of the most basic data we have about the case study, merely for the sake of avoiding a potential counterexample. The second route would be to insist on a different choice of *variables* for the causal model. This response may initially seem just as unmotivated and *ad hoc* as the previous one, but we should not be so quick to dismiss it. In particular, it may be argued that the causal model employed above is impoverished in that it doesn't take into account the fact that Billy's dog catches the ball *before* it hits the partition; hence the temporal aspect of the case may be considered an essential element of the situation that we need to be able to capture.

The problem with this suggestion, however, is that it is hard to see how incorporating the temporal aspect of the situation under consideration would amount to any improvement. For suppose that we introduce two further variables in the original model, 'DM', standing for the ball is in the dog's mouth, and 'BP', standing for the ball hits the partition. To capture the thought that the dog's catch occurs before the ball could hit the partition, we then set up the equations as follows: [BC=1 iff both BP=0 and DM=0]; [DM=1 iff DC=1]; [BP=1 iff PU=1 and DC=0]. We now have a model that allows us to talk about the temporal aspect of the situation under study, in so far as it captures the idea that

the ball would hit the partition only if Billy's dog doesn't catch it. The problem with this model, unfortunately, is that, if the normality ranking stay the same, it makes exactly the same predictions about PU as the previous model. To see this, notice that both PU=1 and BC=0 are actual events, so AC1 holds. Moreover it is true that there is a more normal world where PU=0 and BC=1, namely the world where DC=0; such a world is a witness for the counterfactual: if PU=0 and DC=0, then BC=1. Thus Ac2a holds. Finally, it is true that if PU=1 then BC=0 no matter what value DC has. So Ac2b holds. Since no subset of the variables involved satisfies both AC1 and AC2, the HPN definition still rules PU=1 as a cause of BC=0.

It is not obvious to me that there is a clear fix to this problem, even though I must admit I only have considered the few hypotheses that I find most plausible. My aim in this paragraph was to put into question the idea that default-relativity really is the panacea for every instance of isomorphism, and I hope that by this time the reader has also started doubting that this is the case. In the final two sections of this paper, I want to make a further step in the dialectic, and argue that the problem of isomorphism must be treated very differently from the way HPN theorists have proposed.

7. Doing without Default?

In the previous sections, I have raised some problems for the idea that the structural equations framework must be supplemented with a new function, a normality ordering, reflecting the distinction between default or normal states of affairs and deviant or abnormal ones. If the arguments I have provided are correct, then we have some reason to think that default-relativity does not represent an adequate solution to the problem of isomorphism. My aim in what follows is to move a step further in this debate, and ask where the problems with the HPN proposal leave us in terms of the prospects for a theory of singular causation couched in the language of structural equations. While perhaps some readers might be inclined to think that, with the necessary amendments, some relative of HPN will be able to overcome the challenges I have presented in the previous sections, my view is that the difficulties faced by the HPN proposal are symptomatic of a deeper problem with the structural

equations approach. By way of arguing for this claim, it will be helpful to begin by considering a recent defense of the structural equations approach, by Blanchard and Schaffer (2015).

Blanchard and Schaffer endorse the project of providing an account of singular causation in the language of structural equations; at the same time, they express concerns about the default-relativity solution to the problem of isomorphism. Part of their dissatisfaction with this solution stems from the idea that default and normality reflect cognitive biases in our ordinary reasoning that ought not to enter our theory of causation. As they put it:

> [W]e think that care must be taken to distinguish between those intuitions arising from our competence with the specific concept of actual causation, and those intuitions arising merely from general background biases of cognitive performance. It is a mistake to try to capture intuitions of the latter sort within an account of causation itself (just as it would be a mistake, on noting [similar] effects on probability judgments, to try to incorporate the notions of default and deviant into the probability calculus itself). (p.1)

In effect, Blanchard and Schaffer do not deny that competent speakers might demonstrate all sorts of cognitive biases in their causal judgments, perhaps by refraining to call an event that they consider to be normal *a cause* of another event; still, in their view it would be a mistake to build these sorts of cognitive biases into a theory of causation itself. In the specific case of providing an account of singular causation in the language of structural equations, Blanchard and Schaffer worry that the introduction of default-relativity:

> often seems to us to come close to a free parameter in an otherwise so precise and objectively constrained formalism. (p. 14) ...[T]hings go much more smoothly if we don't have to bother with default-relativity. [...] If we have to complicate the mathematics to add a device for tracking default versus deviant status, then we have compromised this smooth and elegant treatment, and entered a realm where various unclear choices have to be made to even put a causal model on the table (choices that

> moreover just don't seem to matter in the end). All else equal, such complicating and under-constrained unclarities should be avoided if they can. (p.17)

As the authors recognize, doing without default and normality means they must be able to provide an alternative solution to the problem of isomorphism. Interestingly, their strategy for defending the structural equations framework consists in denying that there are ever problematic cases of isomorphism. For instance, when faced with the troublesome case of Assassin-Bodyguard, Blanchard and Schaffer argue that we should pay more attention to the way we actually describe the causal model for that situation, and when we do so it turns out that the resulting model is not isomorphic to the model for Overdetermination. To recall, the causal model initially proposed for Assassin-Bodyguard consisted in the variables: AA (standing for "Assassin does not put the poison"), BA (="Bodyguard puts the antidote") and VS (= "Victim survives"), governed by the structural equation: [VS=1 if either AA=1 or BA=1]. As we have seen in section four, both Hall (2007) and Halpern and Hitchcock (2010) agree that this model constitutes a fair representation of the Assassin-Bodyguard case, and it is precisely for this reason that they take its isomorphism with the model for Overdetermination to be indicative that there must be more to causation than just the structural equations.

On the contrary, Blanchard and Schaffer claim that the proposed model for the Assassin-Bodyguard case is impoverished and must therefore be abandoned. In particular, they claim that the model cannot take into account an essential element of the situation under study, namely whether or not the antidote ever neutralized any poison. In their view, this element is important because if Bodyguard's antidote did not neutralize any poison, then it seems clear that his intervention would not be considered a cause of Victim's survival. Thus they propose to introduce a further variable, NT (="Neutralization occurs"), intended to represent what they take to be the missing element in the original model. The structural equations are then written down as follows: [NT= 1 if both AA=0 and BA=1]; [VS=1 if either AA=0 or NT=1]. In their paper, Blanchard and Schaffer demonstrate that their alternative model is not isomorphic to the model for the Overdetermination case and it also respects the intuitive judgment that the action of Bodyguard putting the antidote in Victim's drink was not a

cause of Victim's survival. Hence they claim that the problem of isomorphism can be overcome by reflecting more carefully on what ought to count as an apt choice of variables for a given situation.

Some worries may be raised against this strategy. Perhaps the most important one is that it is difficult to assess the proposal without having been given a fuller account of what are the appropriate considerations for the modeler to make when choosing a particular set of variables. As pointed out in section three, while defenders of the structural equations approach have provided some rules and guidelines for selecting appropriate variables for a SEF causal model, they are still far away from achieving a formally rigorous account. Indeed, it is precisely because they suspect that no significant progress can be made with regards to specifying the aptness conditions on variables that philosophers like Halpern and Hitchcock are drawn to introduce normality into the picture. Presumably, the lack of a fuller account is also the reason why Blanchard and Schaffer themselves present their view not as an alternative theory, but as a "heuristic" to avoid Hall-style counterexamples, arguing that when cases of isomorphism come up, one should suspect that at least one of the causal models involved is inadequate. However, that's not the same as saying that there aren't problematic cases of isomorphism, and so the worry remains that, if there are such cases, Blanchard and Schaffer would need to resort to the very same complications they are trying to avoid.

I don't mean to raise these problems as in any way decisive against Blanchard's and Schaffer's proposal. It is evident that we are at such an early stage with the project of giving an account of causation within a mathematically rigorous framework that it's unreasonable to ask for anything more than mere guidelines for further research. My reason for mentioning Blanchard's and Schaffer's view is mostly polemical. I see the debate between Schaffer and Blanchard, on one hand, and HPN theorists, on the other, as being representative of a certain way of thinking about the problem of isomorphism for SEF accounts of singular causation, a way of thinking that, despite the superficial differences, these authors in fact share. My worry is that this way of thinking may be questionable in itself, and therefore that in framing the issue as being ultimately one about whether or not we should include default and normality in our formal framework, we may be missing some important connections and failing to ask the right questions.

To see the point I am trying to make, let me start by clarifying what is involved in giving a theory of singular causation. It is a fact about our world that the vast majority of events that actually occur come into being as the result of the concomitance of a number of different factors. For instance, the dropping of a lighted match and the presence of oxygen are among the factors responsible for a forest fire. However, as we have seen, in most occasions when judging matters of causality people focus on only one or a few factors and regard them as causes of an event. For example, people are unlikely to call the presence of the oxygen in the surroundings *a cause* of the forest fire. Given the frequency and systematicity of our practices of causal selection, part of the job of a theory of singular causation is to provide some insight into the mechanisms that underlie these practices. This is not to say that it must be the job of the *philosopher* to provide an account of this sort. For example, Lewis (1973) claims that the principles that govern our practices of causal selection are a matter of pragmatics; if that is true, discovery of these principles may lie outside of the competence of the philosopher. However, it is clearly a requirement upon a philosophical theory that the theory makes it reasonable to believe that some account of our causally selective mechanisms could be given in principle.

It is now customary among philosophers to distinguish between roughly two kinds of causally selective processes (cf. Franklin-Hall 2015). *Vertical* causal selection concerns the level of detail or fine-grainedness used to describe the causal story for a particular event. For instance, describing the causal story leading to a forest fire usually requires mentioning the dropping of a lighted match alone; however, depending on our interests, we might include more details in this story, such as perhaps the height from which the match was dropped, the velocity at which the match fell on the ground, the angle when it first touched the ground, the precise number of molecules of oxygen present in the surroundings at that particular time, etc. *Horizontal* causal selection instead concerns where we draw the line between the cause or causes of an event, on one hand, and the so-called "background" and "enabling" conditions, on the other. For instance, even though the presence of oxygen is certainly one of the factors contributing to the forest fire, it is often disregarded in most ordinary contexts as a cause of the forest fire. Thus, horizontal selection is the process by virtue of which some of the causally relevant factors are singled out among the others and referred to as the *causes* of a certain event.

Thus, any remotely adequate theory of singular causation must respond to the question of what explains or accounts for our causally selective practices, both vertical and horizontal. To the extent that one can find a consensus view on this issue among supporters of the structural equations approach, it seems fair to say that the answer that comes out of their account of causation is a fairly deflationary one: in light of the apparent context-sensitivity of causal claims, there may not be any deep facts about what counts as "the" correct representation of a certain state of affairs, but there are certainly instances of agreements and disagreements among competent speakers about matters of causality, and on the basis of those one can in turn explain the rules for deciding which representation of the causally relevant factors is more appropriate in any given circumstances. Thus, according to defenders of the structural equations approach, the problem of explaining causal selection is at least partly reduced to a problem about what rules are in place for settling disagreements and for the rational critique of alternative causal models.

From the perspective of a defender of the structural equations approach, the problem of isomorphism raised by Hiddleston (2005) and Hall (2007) may be perceived as posing a totally separate kind of question: should we think, following Halpern and Hitchcock, that considerations of default and normality inform our causal judgments, or should we instead follow Blanchard and Schaffer in thinking that the structural equations are enough to account for our causal claims? However, a question that is seldom raised in this context is the following: if it's true that information about default and normality have some sort of influence on our causal judgments, whether as determinants of causal judgments as supported by advocates of HPN or merely as cognitive biases as Blanchard and Schaffer suggest, what is the function of these notions in our causal thinking? In other words, what would be the point or the purpose of having a concept of singular causation that is in some sense sensitive to considerations of default, normality or typicality, as suggested by Assassin-Bodyguard type cases, instead of having one that is insensitive to this peculiar and fairly miscellaneous sort of information?

I believe we should focus on this aspect of the problem brought up by Hiddleston and Hall if we are to make progress understanding the source of the isomorphisms. My aim of the next section is to show that, once a plausible answer is given to the question why default enters our causal thinking in

the way it does, it becomes possible to explain why neither HPN nor any other theory couched in the language of structural equations, including Blanchard's and Schaffer's, can be taken to provide a satisfactory account of singular causation.

8. Explaining Default

What is the function of default and normality in our causal thinking? Psychologists Kahneman and Miller (1986) once observed that "an event is more likely to be undone by altering exceptional than routine aspects of the causal chain that led to it" (p. 143). This observation may seem initially puzzling because "exceptional" and "routine" are in large part subjective and human-centered notions, and so it may strike us as a lucky coincidence that they would find such a neat application to the discovery of the world's causal structure. But the puzzlement disappears once we realize that what counts as "exceptional" and "routine" may be, as Woodward (2007) puts it, "as much a *product* of our practices of causal judgment (or our practical and theoretical interests) as an independent input to them" (p.21). In other words, it is plausible that, in the process of acting and manipulating things in the world, we form a conception of what is exceptional and routine, a conception that, as Kahneman and Miller suggest, can in turn be used to make our capacity for causal judgment quicker and more reliable.

Therefore, Kahneman and Miller's observation is best interpreted as bringing out an interesting interconnection between our capacity to discriminate between default and deviant states of affairs, on one hand, and our capacity for causal thinking, on the other. The question in front of us then becomes how best to understand this interconnection. Although philosophers might disagree on this issue, one plausible answer begins with the theoretical and practical interests connected with the assessment of singular causal claims. Arguably, our concept of singular causation is a fairly specialized tool, at least one purpose of which is to allow us to articulate information that is relevant to an efficient use of our capacities for agency and manipulation, as well as for attributing responsibility and blame among members of our community. Causal selection plays an important role in achieving these aims, as it

allows us to focus our attention on the factors that are most appropriate as targets of intervention or for the attribution of responsibility. For instance, by appropriately selecting the lighting of the match as *the* cause of the forest fire, we are in a position to make inferences about what or who is responsible and ought to be blamed for the event in question.

Given that default and normality appear to have some considerable influence on our causal thinking, it is natural to suppose that possessing information about what counts as default and what doesn't could be of some help in achieving the aims or fulfilling the functions of our concept of singular causation. What kind of help? One hypothesis, which I find very suggestive, is that the reason why default and normality are connected to causal thinking must lie with the fact that they help us with the making of the horizontal selection between causes and merely enabling conditions; in turn, this is valuable because the making of this distinction helps us with attributing responsibility and blame as well as finding appropriate target of intervention. This view finds intuitive support in many of the examples already discussed. For instance, knowing that the presence of oxygen in the surroundings is normal rather than abnormal might help us focus our attention on the lighting of the match as the most relevant factor contributing to, and indeed causing, the forest fire. Similarly, knowing the normal functions of the Queen might help us focus our attention on the more relevant fact of gardener's failure to water the plants, which caused their death.

These examples suggest that possessing certain information about default and normality is helpful for the purpose of horizontally selecting between causes and merely enabling conditions. Of course, at this point one would also need to provide a story about what makes the lighting of the match or the gardener's omission relevant factors in the first place. Although what I will give below is nothing but a sketch of a view that I find plausible, it is important to signal that this is an issue on which a lot of philosophical work has been done in recent years, and so there may be more than one option available. At any rate, a promising approach begins with the idea that the causally relevant factors are those that *make a difference* to the occurrence of a certain outcome. Lewis (1973) introduced this idea in a famous article on causation. He wrote:

> We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects—some of them, at least, and usually all—would have been absent as well. (p. 160)

Following Lewis, we can then give an initial characterization of the notion of 'making a difference' in counterfactual terms: *c* makes a difference to *e* just in case, if *c* had not occurred, *e* would not have occurred. For instance, one could say that the lighting of the match made a difference to the forest fire because, had the match not been lighted, the forest fire wouldn't have occurred. Of course, one could say the same about the presence of oxygen in the atmosphere: had there been no oxygen at the time the match was lighted, there would have been no fire. But this is not a problem for the view I am defending. What difference-making does is merely giving an account of what it means for some event to be causally relevant. The definition of singular causation is an entirely different issue. As I said earlier, a theory of singular causation would need to say something about the way we select, among the various causally relevant factors, those that we deem as "the" causes of a certain event. The proposal here is simply that difference making is sufficient for some event to be a causally relevant factor; such an account was given only with the aim of supporting the hypothesis that default and normality might play a role in causal selection, by allowing us to make quick and reliable decisions about which of the many causally relevant factors should count as *the* cause of a particular event.

In sum, I believe that there is some plausibility to the idea that information about default and normality play a role in horizontal causal selection, by helping us with the tracing of the distinction between causes and merely enabling conditions. Even though what I have given is only a sketch of a proposal, the hypothesis that it is based on seems to provide a nice explanation why our concept of singular causation would be sensitive to the miscellaneous sort of information that falls under the notions of default and normality. In a nutshell, the answer is that default and normality contribute to horizontal causal selection; in turn, horizontal causal selection promotes the theoretical and practical interests connected with the attribution of responsibility and blame, as well as the discovery of appropriate targets of intervention. For instance, while both the lighting of the match and the presence

of oxygen are relevant factors in the causal story that leads to the forest fire (in the sense of 'relevant factors' that I have tried to spell out before through the notion of difference-making), by relying on what is "default" or "normal" in those circumstances we can zero in on the lighting of the match and deem it *the* cause of the forest fire. I have suggested that doing so may serve the theoretical and practical interests connected with causal selection, including but not limited to the potential social benefits derived from the attribution of responsibility and blame.

The relative plausibility of the proposed hypothesis raises more than one problem for SEF theorists. Perhaps the clearest and most immediate one is that any attempt to trace a distinction between what SEF theorists call "the problem of the variables" and "the problem of isomorphism" is doomed to failure. For if it is true that background assumptions about default and normality help us by solving problems of horizontal causal selection, then it is very plausible that those assumptions figure in our decisions about which events we take to be relevant in a particular situation. An observation that support this conclusion is that, without a normality ordering on possible worlds, one can generate cases of isomorphism simply by introducing a variable in a causal model standing for an event in its default or normal state. For instance, suppose that our causal model for the Gardener case is given by the variables: 'GD', standing for "Gardener doesn't water the plants", and 'PD', standing for "plants die", with equations: [PD=1 iff GD=1]. Clearly, by adding a variable for the Queen's behavior in this model, say 'QD', for "Queen doesn't water the plants", we generate an instance of isomorphism with the model M_0 of section two. To avoid the isomorphism, one might alternatively assign a normality ranking to the events in the model, or simply decide to exclude 'QD' as irrelevant. Either way, background assumptions about default and normality must be employed at some stage in the selection of the appropriate model, which shows that there aren't really two problems here, one concerning the appropriate selection of the variables for the model and the other concerning how to avoid the isomorphism. Really the only problem is how to make sense of our practices of causal judgment in light of the fact that default and normality appear to play a role in causal selection.

We therefore come to a first important criticism of the diagnosis of the problem of isomorphism provided by HPN theorists. On the proposals by Halpern (2008) and Halpern and Hitchcock (2009), we are supposed to infer from the few cases of isomorphism pointed out by Hiddleston (2005) and

Hall (2007) that our causal models must contain some sort of reference to a particular standard of normality or typicality. In the previous sections, I have argued that this hypothesis of default-relativity imposes too heavy a burden on our practices: there are far too many cases in which it seems clear that none of the considerations concerning normality and default have anything at all to do with our judgments. The hypothesis that I am advancing in this section helps us see why this is the case. On my hypothesis, default and normality enter into our causal judgments in the way of solving problems of horizontal selection. It follows that, in cases where the selection has already been made, it is not only possible but also likely that there be no role for considerations of normality and default to play. For instance, already in the way of describing the cases of Train and Engineers, I made no mention of weather conditions or temperature; details about the location, physical and psychological conditions of the engineers were excluded from the story; and so on. All these aspects of the situation were assumed to be "normal". Thus, when I say that in these cases the relevant kind of causal selection has already been made, I mean that in these cases background assumptions about default and normality have entered the work of "parsing" between causally relevant factors at some earlier stage; no wonder, then, that by the time we have cut down the causal structure to only a few events, be they the switch pointing one way or another, or the removal of blocks by way of the activation of a security system, we would find no other way but to literally make up the normality ordering for those events in such a way as to have the formal theory accord with our judgment.

Furthermore, from the vantage point of my hypothesis it is to be expected that there would be cases, such as the case of preventative preemption discussed in section six, in which our intuitions about what counts as normal and default would be hard to square with the aim of building up a model that predicts our intuitive causal judgment. Because information about default and normality has to be introduced in the formalism even though it plays no role in our causal judgments, it is not unlikely that at least in some cases the introduction of a normality ranking would be in effect detrimental to the scope of getting at the right verdict for the situation under study. For instance, in the case of preventative preemption of section six, noting that the presence of a transparent partition in a field is a very abnormal event given the circumstances misleads us into building a causal model that makes the wrong predictions about our causal judgments. Thus, I believe that the hypothesis that default and

normality enter into the mechanisms of horizontal causal selection gives a nice and powerful explanation of why the problems I have raised earlier for defenders of HPN are not in any way accidental, but are instead predictable consequences of taking this sort of approach. What this suggests, in my view, is that the way in which default and normality enter into our causal thinking is not adequately captured by the default-relativity solution to the problem of isomorphism, which in effect completely mistakes the sense in which our causal judgments are sensitive to information about default.

At this point, a defender of the structural equations approach to singular causation may argue that, even if successful, my arguments above only really work against one specific proposal for solving the problem of isomorphism. In other words, they may concede that the introduction of a normality ordering is not the right kind of tool for solving the problem of isomorphism; at the same time, however, they may insist that there is no reason to doubt that an alternative solution to the problem may be found that aligns with the spirit of the structural equations approach. One such alternative may be Blanchard's and Schaffer's account that was briefly reviewed in section seven; perhaps other alternatives will become available as research in this area increases. By way of concluding this paper, I want to express my skepticism about the prospects for this last defense of the structural equations approach.

It seems to me that the hypothesis that the influence of default and normality must ultimately be explained by the practical and theoretical interests connected with the correct assessment of causal claims allows us to bring out a crucial problem with the deflationary view of causal selection supported by all SEF theorists. To recall, this deflationary view states that there are no any deep facts about what counts as "the" correct representation of a given situation; there are just cases of agreement and disagreement about matters of causality, and by reflecting on those we can then formulate a set of regulative principles for the rational critique of alternative causal models. However, it is arguable that, if indeed true, the hypothesis I have put forward about the role of default in our causal thinking would go a long way towards providing the basic elements of an explanation of our practices of causal judgment, by connecting them to the practical and theoretical advantages brought to us by the capacity for causal selection. On the basis of this connection one might, for instance,

explain a person's judgment that event *e* was a singular cause of event *c* at least partly in terms of the practical and theoretical interests that are involved in making this judgment in the particular circumstances or context of evaluation. Unfortunately for SEF theorists, however, this seems to be precisely the kind of deep and unifying explanation of our practices of agreement and disagreement that their deflationary account says not to exist.

The significance of this point cannot be stressed enough. By elaborating on a claim that both HPN theorists and Blanchard and Schaffer agree on, namely that our singular causal claims are influenced by background information concerning what's deviant and what's default, I have argued that a plausible explanation why our concept of singular causation is sensitive to this sort of influence must have something to do with the fact that the default/deviant distinction helps us solve problems of horizontal causal selection. However, what we have on the table now, if we take this hypothesis seriously, is the beginning of a story that connects our practices of causal judgment to the practical and theoretical interests associated with horizontal causal selection and, more generally, with the correct assessment of singular causal claims. In light of this connection, it is plausible to suppose that a more informative account of singular causal claims can be found that places more substantive constraints on our practices of causal judgment, constraints that limit the range of what can count as an apt representation of the causal structure of a given situation significantly more than what defenders of the structural equations approach can countenance. This alternative account would make it possible to understand questions about the relative aptness of alternative causal models for a given situation at least in part as normative questions concerning the extent to which representing the causal structure of a given situation in one way rather than another helps us pursue the practical and theoretical interests that are associated with causal selection. Relatedly, such a theory would allow us to evaluate the proposed rules for the rational critique of alternative models not just on the basis of how close they come to predict our intuitive causal judgments in each case, but more importantly on the basis of how efficiently they allow us to pursue the practical and theoretical interests that we have for correctly assessing singular causal claims.

In sum, I believe that, if we take the hypothesis that I presented above seriously, we have a reason to think that the deflationary account supported by SEF theorists does not provide the full story

concerning our practices of causal judgments. The reason for this is that, merely by reflecting on the role of default and normality in our causal thinking, we seem to have reached some suitable ground to explain our practices of singular causal judgments, not just in terms of which models or standards of normality competent speakers are implicitly working with in some given circumstances, as SEF theorists propose to do, but ultimately in terms of the sorts of practical and theoretical interests that underlie our practices of causal selection. This conclusion stands in opposition to, and indeed refutes, the deflationary stance proposed by SEF theorists concerning the status of our representations of causal structures. For while on the deflationary account the explanation of our practices of causal judgment bottoms out, so to speak, in the facts about agreement and disagreement among competent speakers, on the alternative approach that I have suggested in this section our practices of causal judgment are responsive to certain normative constraints that are in place as the result of the fact that not all representations of a given situation are equally well suited to pursue the practical and theoretical aims connected to our practices of causal selection. The recognition that such normative constraints are in place on our practices allows us to see the inherent limitations of the structural equations framework as a basis for a theory of singular causation. This does not mean that the formal tools provided by SEF cannot be useful, in many circumstances, as a way of modeling the causal structure of a given situation; it only implies that it is a mistake to think we can turn the formal tools of representation provided by the structural equations framework into a full-fledged philosophical theory of singular causation.

REFERENCES

Blanchard, T. and Schaffer, J. (2015). "Cause without Default". To appear in Beebee, H., Hitchcock, C. and Price, H. *Making a Difference*. Oxford: Oxford University Press.

Collingwood, R. (1940): *An Essay on Metaphysics*. Oxford: Clarendon Press.

Collins, J., Hall, N. and Paul, L.A. (2004). *Causation and Counterfactuals*. Cambridge: MIT Press.

Cushman F., Knobe, J. and Sinnott-Armstrong, W. (2008). "Moral appraisals affect doing/allowing judgments". *Cognition* 108 (1): 281-289.

Cushman, F. (2009). "The role of moral judgment in causal and intentional attribution: What we say or how we think?". Unpublished manuscript

Franklin-Hall, L. (2015). "Explaining causal selection with explanatory causal economy: Biology and beyond". In C. Malaterre and P.-A. Braillard (eds.), *How Does Biology Explain?* Springer-Verlag, Heidelberg.

Gallow, J. (2015). *The Emergence of Causation*. PhD Thesis

Hall, N. (2007). "Structural Equations and Causation". *Philosophical Studies*. 132: 109-136.

Halpern, J. Y and Hitchcock, C. (2015). "Graded causation and defaults". To appear in the *British Journal for the Philosophy of Science*.

Halpern, J. Y. and Pearl, J. (2005). "Causes and explanations: A structural-model approach. Part I: Causes". *British Journal for Philosophy of Science*. 56(4): 843-887.

Halpern, J. Y. (2008). "Defaults and normality in causal structures". In *Principles of Knowledge Representation and Reasoning: Proc. Eleventh International Conference (KR '08).* 198-208.

Halpern, J. Y. and Hitchcock, C. (2010). "Actual Causation and the Art of Modeling". In R. Dechter, H. Geffner, and J.Y. Halpern, editors, *Causality, Probability, and Heuristics: A Tribute to Judea Pearl*, pp. 383–406. College Publications, London, 2010.

Hiddleston, E. (2005). "Causal Powers". *British Journal for Philosophy of Science*. 56: 27-59.

Hitchcock, C. (2007). "Prevention, preemption and the principle of sufficient reason". *Philosophical Review*. 116: 495-532.

Hitchcock, C. and Knobe, J. (2009). “Cause and Norm”. *Journal of Philosophy*. 106 (11): 587-612.

Kahneman, D. and Miller, D. (1986). “Norm Theory: Comparing Reality to its Alternatives”. Psychological Review 93: 136-153.

Lewis, D. (1973) “Causation”. *Journal of Philosophy,* 70:556–567. Reprinted with added “Postscripts” in Lewis, D. (1986). *Philosophical Papers Volume II*. Oxford University Press. pp. 159–213.

Maudlin, T. (2004). “Causation, Counterfactuals and the Third Factor”. In Collins, J., Hall, N. and Paul, L.A. (2004).

McDermott, M. (1995). “Redundant Causation”. *British Journal of Philosophy of Science*. 46: 523-544.

Menzies, P. (2004). “Difference-Making in Context”. In Collins, J., Hall, N. and Paul, L.A. (2004).

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

Woodward, J. (2007). “Sensitive and Insensitive Causation”. Philosophical Review. 115 (1): 1-50.