

Normal Causes for Normal Effects: Reinvigorating the Correspondence Hypothesis About Judgments of Actual Causation

Totte Harinen¹ 

Received: 24 November 2015 / Accepted: 13 January 2017 / Published online: 10 July 2017
© The Author(s) 2017. This article is an open access publication

Abstract There have been several recent attempts to model ordinary intuitions about actual causation by combining a counterfactual definition of the causal relation with an abnormality-based account of causal judgments. In these models, the underlying psychological theory is that people automatically focus on *abnormal* events when judging the actual causes of an effect. This approach has enabled authors such as Halpern and Hitchcock (Br J Philos Sci axt050, 2014) to capture an impressive array of ordinary causal intuitions. However, in this paper I demonstrate how these abnormality-based accounts still systematically fail to predict ordinary causal judgments in specific types of scenarios: those in which the effect is normal. I will argue that the reason for this is that the underlying psychological theory is wrong: the idea that intuitive actual causes are abnormal events is only partially correct. To model ordinary judgments more realistically, researchers working in this area must adopt a more plausible underlying psychological theory: the *correspondence hypothesis* about judgments of actual causation. One of the consequences of this correspondence hypothesis is that normal effects are judged to have *normal* causes.

1 Introduction

Most recent work on causal judgments builds upon the path-breaking analyses of Kahneman and his colleagues (Kahneman and Tversky 1981; Tversky and Kahneman 1973; Kahneman and Miller 1986) who demonstrated that people tend to pick out *abnormal* events or values of variables as the ‘actual’ or ‘token’ causes when they are presented with a causal scenario. However, as some observers were

✉ Totte Harinen
totte.harinen@kcl.ac.uk

¹ Department of Philosophy, King’s College London, The Strand, London WC2R 2LS, UK

quick to point out, this early work on causation and normality tended to focus almost exclusively on *abnormal effects*, i.e. exceptional, atypical or surprising outcomes of causal processes, such as motorbike accidents (Gavanski and Wells 1989). The critiques argued that this ‘abnormality bias’ (Prentice and Koehler 2002) led to psychological theories about causal judgments in which perceived token causes were falsely assumed to be *always* abnormal. Taking their inspiration from Mill’s (1856) original insights, some of these psychologists (cf. Einhorn and Hogarth 1986; Gavanski and Wells 1989) developed and tested the following alternative hypothesis about causal judgments: the crucial issue is *not* normality or abnormality as such but rather that there is a relation of *representativeness* between causes and effects. One of the predictions following from this ‘correspondence hypothesis’ is that psychological subjects should judge *normal* outcomes to have *normal* causes. Gavanski and Wells (1989) tested this prediction empirically and found it supported. People tend to judge abnormal effects as having abnormal causes and normal effects as having normal causes.

In this paper, I argue that the abnormality bias has resurfaced in recent philosophical and scientific work on actual causation. Following Hart and Honoré’s seminal (1985) work, there have been several recent attempts to account for folk intuitions about actual causes by incorporating normality considerations into counterfactual accounts of causation (cf. Menzies 2007; Hall 2007; Hitchcock 2007; Halpern and Hitchcock 2014). The very basic idea in all these accounts is that, in any causal scenario, ordinary speakers tend to assign a stronger causal status to those factors that are unexpected, surprising, atypical or abnormal. Thus, for example, if we are more likely to say that the occurrence of a forest fire is the result of someone’s lighting a match rather than the result of there being oxygen present in the atmosphere—even though both factors are counterfactual difference-makers for the fire—it is because someone’s lighting a match is an abnormal event whereas there being oxygen present in the atmosphere is very normal indeed. Developing their formal philosophical accounts by building upon this observation, authors such as Halpern and Hitchcock have assumed that intuitive actual causes are *always* abnormal or ‘atypical’ values of variables (I will return to the relationship between ‘normality’ and ‘typicality’ later). In doing so, I argue, these authors have codified the abnormality bias into their conceptual framework.

In my discussion, I will specifically focus on the recent work by Halpern and Hitchcock (2013, 2014), which has provided us with the most sophisticated abnormality-based philosophical treatment of actual causation so far. The method that Halpern and Hitchcock use is that of taking a set of causal scenarios and trying to provide a formal definition of actual causation that agrees with folk judgments about those scenarios. The definition’s agreement with folk intuitions can then be treated as ‘evidence’ for or against it.¹ Now as I emphasized above, Halpern and Hitchcock’s philosophical theory is based on the assumption that folk judgments about actual causation are governed by the following rule: actual causes are always abnormal. It follows, then, that their definition of actual causation implies that

¹ Legitimate worries can be raised as to the usefulness of that approach (cf. Glymour et al. 2010; Danks 2013). But in this essay I want to put these methodological objections to one side and discuss some causal scenarios that normality-based philosophical accounts do not get right.

normal effects are judged to have abnormal causes. This is the opposite of what the correspondence hypothesis predicts, which is that normal effects are judged to have normal causes. Therefore, the correspondence hypothesis implies that Halpern and Hitchcock's account of actual causation disagrees with folk intuitions when the effect in a causal scenario is normal. Below, I will demonstrate that this is indeed the case. Further, I will argue that empirical and theoretical considerations suggest that the correspondence hypothesis is superior to its abnormality-based alternatives. Formal accounts of actual causation should therefore be compatible with the correspondence hypothesis.

The paper is divided into two main parts. The first part consists of Sects. 2 and 3, where I discuss Halpern and Hitchcock's theory of actual causation as a sophisticated abnormality-based approach. I start, in Sect. 2, by introducing the Halpern–Pearl formalism for actual causation, which has been presented in various places over the last 10 years (cf. Halpern and Hitchcock 2011, 2014; Halpern and Pearl 2005; Halpern 2013, 2016). I will then explain how Halpern and Hitchcock build abnormality considerations into that formalism to account for folk intuitions. Then, in Sect. 3, I will proceed to discuss several examples of causal scenarios that Halpern and Hitchcock's account does and does not get right. I will particularly focus on certain *prevention* cases in which the value of the effect variable is intuitively *normal*. I will demonstrate that Halpern and Hitchcock's account does not deliver the intuitively correct result when applied to these examples. Section 4, then, marks the beginning of the second part of the paper. Here, I begin by describing the exact way in which Halpern and Hitchcock are committed to the idea that actual causes are abnormal. I will then describe the correspondence hypothesis and show how Halpern and Hitchcock's account is in tension with it. I will further argue that the correspondence hypothesis is better supported by empirical evidence compared to its major abnormality-based alternative. To close the paper, I will then discuss how causes can influence the normality of effects and the dynamics in which this results in certain prevention scenarios.

2 The Halpern–Pearl Formalism

We start with a causal model M , which is a tuple $(\mathcal{U}, \mathcal{V}, \mathcal{E})$, where \mathcal{U} is a set of exogenous variables, \mathcal{V} is a set of endogenous variables, and \mathcal{E} is a set of structural equations relating the values of the variables. It is usually helpful to assume that M is acyclic, which intuitively means that the model does not contain feedback loops between variables. More formally, in an acyclic causal model there is some total ordering \prec such that, if $X \prec Y$, then the value of X may (or may not) influence the value of Y while the value of Y cannot influence the value of X .

In the definition of actual causation below, ϕ denotes a Boolean combination of 'primitive events' of the form $X = x$ where $X \in \mathcal{V}$ and x is among the possible values of X . The abbreviation $[\vec{Y} \leftarrow \vec{y}]$ denotes the case in which variables $Y_1, \dots, Y_k \in \mathcal{V}$ are set to some of their possible values y_i for $i = 1, \dots, k$. If the value of Y_i is set to one of its possible values in this way, the result is a new causal model which is otherwise identical to M except that the structural equation for Y_i is

replaced with $Y_i = y_i$. $[\vec{Y} \leftarrow \vec{y}]\phi$ states that ϕ would hold if each Y_i were to be set to a possible value y_i . Such a ‘causal formula’, abbreviated ψ , requires some *context* \vec{u} , which is the assignment of the values of the exogenous variables in the model. The fact that some causal formula holds in some causal model and context can then be expressed as $(M, \vec{u}) \models \psi$.²

We can now state the Halpern-Pearl definition of actual causation. $\vec{X} = \vec{x}$ counts as an actual cause of ϕ in (M, \vec{u}) if the following three conditions hold (quoted from Halpern 2013, p. 3, with inconsequential changes in notation):

- AC1. $(M, \vec{u}) \models (\vec{X} = \vec{x})$ and $(M, \vec{u}) \models \phi$
- AC2. There is a partition of \mathcal{V} (the set of endogenous variables) into two subsets \vec{Z} and \vec{W} with $\vec{X} \subseteq \vec{Z}$ and a setting \vec{x}' and \vec{w} of the variables in \vec{X} and \vec{W} , respectively, such that if $(M, \vec{u}) \models Z = z^*$ for all $Z \in \vec{Z}$, then both of the following conditions hold:
 - (a) $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}]\neg\phi$
 - (b) $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}, \vec{W}' \leftarrow \vec{w}, \vec{Z}' \leftarrow \vec{z}^*]\phi$ for all subsets \vec{W}' of \vec{W} and all subsets \vec{Z}' of \vec{Z} , where I abuse notation and write $\vec{W}' \leftarrow \vec{w}$ to denote the assignment where the variables in \vec{W}' get the same values as they would in the assignment $\vec{W} \leftarrow \vec{w}$.
- AC3. \vec{X} is minimal; no subset of \vec{X} satisfies conditions AC1 and AC2.

Here, we say that the tuple $(\vec{W}, \vec{w}, \vec{x}')$ is a *witness* to the fact that $\vec{X} = \vec{x}$ is an actual cause of ϕ .

Condition AC1 requires that $\vec{X} = \vec{x}$ and ϕ actually be the case. This reflects the fact that we are defining the concept of an actual cause. Condition AC3 is the familiar requirement that the cause should not involve elements that are redundant with respect to the effect.

The major theoretical contribution of Halpern and Pearl’s definition lies in AC2. Here, the set of endogenous variables is partitioned into two subsets \vec{Z} and \vec{W} where the variables in \vec{Z} correspond to the causal path from \vec{X} to ϕ . The variables in \vec{W} , again, are those that are not on the path from \vec{X} to ϕ . Now, condition AC2(a) states that changing the value of \vec{X} to some value $\vec{x}' \neq \vec{x}$ results in it being the case that $\neg\phi$ holds. This betrays the fact that the account of causation in play is counterfactual: setting \vec{X} into a non-actual value results in $\neg\phi$ holding, contrary to what is actually the case. However, it need not be true that setting the value of \vec{X} to \vec{x}' *invariably* results in $\neg\phi$. Rather, setting the value of \vec{X} to \vec{x}' need only result in $\neg\phi$ under the *contingency* that the variables in the set of ‘off-path’ variables \vec{W} are fixed to some values \vec{w} . This captures the common idea that a causal dependency between two or more variables sometimes becomes ‘visible’ just in the case some ‘off-path’ variables in the causal model are kept fixed.

² For a more technical exposition, see any of the papers by Halpern cited in the introduction.

Condition AC2(b) imposes some additional constraints on the variables in the two sets. The requirement is that ϕ holds if we set \vec{X} to \vec{x} and (1) any subset of the variables in \vec{W} get the same values as they would in $\vec{W} \leftarrow \vec{w}$ and (2) any subset of the variables in \vec{Z} get the same values as they do in the *actual* world. The point of this requirement is to ensure that even if fixing the values of the variables in \vec{W}' changes the values of the variables in \vec{Z} , this change is not what makes it the case that ϕ holds when $\vec{X} = \vec{x}$. To capture this, the relationship between $\vec{X} = \vec{x}$ and ϕ is required to hold even when a subset of the variables in \vec{Z} are set in their actual values, i.e. the values implied by (M, \vec{u}) .

The ‘witness’ for the causal relation between $\vec{X} = \vec{x}$ and ϕ is the setting of variables in the relevant model that we consult in order to reveal the counterfactual dependency between the two. Importantly, such witnesses can be ranked in terms of their normality. Halpern and Hitchcock (2014) utilize a partial preorder \succeq to do so. This relation is transitive and reflexive but not total.

Take two ‘small worlds’, s and s' , which are complete specifications of the endogenous variables in a causal model. If s is strictly more normal than s' , then Halpern and Hitchcock write $s \succeq s'$. If $s \succeq s'$ and $s' \succeq s$, then they write $s \equiv s'$. In order for s and s' to be comparable, there must be some nonempty set \vec{X} of variables that take more typical values in one of the worlds and less typical values in the other. However, if it is the case that there are two sets of variables \vec{X} and \vec{Y} such that the variables in \vec{X} take more typical values in one of the worlds but the variables in \vec{Y} take more typical values in the other, then s and s' are incomparable. This is what is meant by saying that \succeq is not total.

Let $s_{\vec{u}}$ denote the actual small world and $s_{\vec{w}, \vec{x}, \vec{u}}$ the world in which \vec{W} is set to \vec{w} , \vec{X} is set to \vec{x} and the context is \vec{u} . Then Halpern and Hitchcock require that, in order for $\vec{X} = \vec{x}$ to be an actual cause of ϕ , conditions AC1 thru AC3 must be satisfied, and additionally in AC2(a) it must be the case that $s_{\vec{w}, \vec{x}, \vec{u}} \succeq s_{\vec{u}}$. In other words, the small world that is the witness for the causal relation between $\vec{X} = \vec{x}$ and ϕ must be at least as normal as the actual world. If there are many witnesses, it is possible to order them according to their normality. Call the most normal world for $\vec{X} = \vec{x}$ being a cause of ϕ as the ‘best witness’. Halpern and Hitchcock conjecture that, in situations where there are multiple causes for some ϕ , people tend to pick out those that have the best witnesses as the actual causes. In what follows, whenever I talk about witnesses, I talk about *small witness worlds*.

3 Examples

The point of combining normality considerations with a counterfactual definition of actual causation is that doing so enables us to account for a variety of old causal chestnuts with which counterfactual definitions *alone* struggle. These chestnuts include Prevention, Double Prevention, Bogus Prevention, Pre-emption, Omission,

Selection, Switching, Emphasis, Transitivity and Thrombosis. The common factor in all these cases is that the counterfactual relationship between two variables is usually *masked* in one way or another. This is why we need to consider a contingency—a setting of the off-path variables in our model—that reveals those masked counterfactual relationships. This is what definitions of actual causation such as the H-P account described above deliver. But even after we have successfully specified the contingency, we are still left with some cases where the H-P definition delivers the intuitively wrong result. These are the cases that we hope to be able to get right by adding the normality ranking into our theory of actual causation.

In this section, I will discuss some causal scenarios with which Halpern and Hitchcock's normality-based account struggles. I want to emphasize that there are also several cases, which I have no space to discuss, that their account handles elegantly (for which see Halpern and Hitchcock 2014). What is common between the cases that Halpern and Hitchcock's account gets right is that in all of them the actual value of the effect variable is abnormal. That is why their account manages to pick out those effects' intuitively correct, abnormal actual causes. Conversely, the problematic examples that I will discuss in this section have the common feature that the value of the effect variable is *normal*.

Before proceeding any further, let me note a terminological difference between my treatment and Halpern and Hitchcock's. When Halpern and Hitchcock talk about what I call the 'normality' of the value of a variable, they use the term 'default' or 'typical'. This is, in their own words, an 'arbitrary stipulation' made for 'ease of exposition' (Halpern and Hitchcock 2014, p. 434). In what follows, I will *not* adopt Halpern and Hitchcock's terminology. The reason is that the distinction between the 'normality' of worlds versus the 'typicality' of values is not present in the psychological literature of which Halpern and Hitchcock's account is intended to be a formalization (more on this later). Because it is one of my main goals in this paper to bring that literature to bear, it is not helpful to introduce said arbitrary distinction to the discussion. Consequently, I will from now on use the term 'normal' in such a way that it also covers the notions 'default' and 'typical'. In Sect. 4, I will explain in detail how *normality* and related notions are understood in the relevant psychological literature.

With this conceptual clarification in mind, let us now examine how Halpern and Hitchcock's account works by discussing the following perfectly ordinary, unproblematic case of pre-emption:

Example 1 Suzy and Bill, both competent rock throwers, are determined to destroy a nearby glass bottle. Suzy throws a rock that hits the bottle and the bottle shatters. Had Suzy not thrown her rock, Bill would have thrown his, in which case the bottle would also have shattered.

Let us model the case with the following variables:

$$\begin{aligned}
 ST &= \begin{cases} 1 & \text{if Suzy throws} \\ 0 & \text{otherwise} \end{cases} \\
 BT &= \begin{cases} 1 & \text{if Bill throws} \\ 0 & \text{otherwise} \end{cases} \\
 BS &= \begin{cases} 1 & \text{if the bottle shatters} \\ 0 & \text{otherwise} \end{cases}
 \end{aligned}$$

With these variables at hand, we can observe that the following equations capture the causal structure of the situation:

$$\begin{aligned}
 ST &= 1 \\
 BT &= 1 - ST \\
 BS &= \max(ST, BT)
 \end{aligned}$$

The H-P definition of actual causation yields the correct result with $\vec{Z} = \{ST, BS\}$, $\vec{W} = \{BT\}$ and the value of the variable in \vec{W} set to 0. The witness for the causal relation between the values of ST and BS is the small world ($ST = 0, BT = 0, BS = 0$). This is the contingency that we need to consider in order to unmask the counterfactual relationship between $ST = 1$ and $BS = 1$. Because that relationship is revealed only when we consider a contingency, it is easy to see that we cannot naively require that the counterfactual relationship between the values of ST and BS must be visible under all conditions in order for there to be causation between them. Finally, note that the contingency that we consider in order to unmask the relationship, namely ($ST = 0, BT = 0, BS = 0$), is intuitively more normal than the actual world ($ST = 1, BT = 0, BS = 1$) is. Thus, $ST = 1$ counts as an actual cause of $BS = 1$ under Halpern and Hitchcock’s account. This is to be expected, given that the actual value of the effect variable in this causal scenario is abnormal.

Moving on to more challenging examples, let us consider a case of bogus prevention:

Example 2 Assassin plans to put a lethal dose of poison in Victim’s coffee. At the very last moment, she changes her mind and refrains. Bodyguard puts effective antidote in the coffee. Victim drinks the coffee and survives.

Here we can use the following variables:

$$\begin{aligned}
 A &= \begin{cases} 1 & \text{if Assassin puts in poison} \\ 0 & \text{otherwise} \end{cases} \\
 B &= \begin{cases} 1 & \text{if Bodyguard puts in antidote} \\ 0 & \text{otherwise} \end{cases} \\
 VS &= \begin{cases} 1 & \text{if Victim survives} \\ 0 & \text{otherwise} \end{cases}
 \end{aligned}$$

We need just one equation to capture the causal structure:

$$VS = \max((1 - A), B)$$

Without normality considerations, the H-P definition counts $B = 1$ as a cause of $VS = 1$ in the actual world where $A = 0$. For if we partition the variables into $\vec{W} = \{A\}$ and $\vec{Z} = \{B, VS\}$, set $A = 1$ and switch $B = 0$, we end up with $VS = 0$. AC2(b) doesn't help because the members of $\vec{Z} - \vec{X}$ consist of $\{B\}$ and VS . But observe now that the world $s_{\vec{w}, \vec{x}, \vec{u}}$ is $(A = 1, B = 0, VS = 0)$ while the world $s_{\vec{u}}$ is $(A = 0, B = 1, VS = 1)$. Assuming that the normal values for the cause variables here are $A = 0$ and $B = 0$, we can now observe that the two worlds are, for Halpern and Hitchcock, *incomparable*. There is a singleton set $\{A\}$ with a more normal value in $s_{\vec{u}}$; and there is the singleton set $\{B\}$ with a more normal value in $s_{\vec{w}, \vec{x}, \vec{u}}$. Given Halpern and Hitchcock's requirement that $s_{\vec{w}, \vec{x}, \vec{u}}$ should be at least as normal as $s_{\vec{u}}$, $B = 1$ does not count as a cause of $VS = 1$, just as required.

Let me here mark an ambiguity in Halpern and Hitchcock's account. For it is not clear whether one should incorporate the value of the *effect* variable in the normality comparison. Certainly, when Halpern and Hitchcock (2014) discuss the several examples that they are able to sort out with their account, they focus on the normality of the cause variables. A world counts as more normal than another if there is a non-empty set of variables that take more normal values in that world, and no variables that take more normal values in the other world. Elsewhere (2013, pp. 17–18) Halpern and Hitchcock suggest that the normal value of the effect variable is determined by the values of the cause variables. If so, then the value of the effect variable can be ignored in the comparison simply because it is always normal. I agree that the normality of the value of the effect variable depends in some ways on the values of the cause variables, although in my view there are causal scenarios, such as prevention, in which it is nevertheless regarded as abnormal. I will return to this in detail later.

Halpern and Hitchcock's solution to bogus prevention strikes me as independently implausible.³ The reason is that I fail to intuit that the two settings of variables—the actual setting and its witness—are impossible to compare in the above example. Consequently, it seems that if Halpern and Hitchcock's account on this occasion picks out the intuitively correct actual cause, it is a matter of coincidence rather than the outcome of a strongly rooted connection between their account and the psychological mechanisms of actual causation attribution. This suggests the following question: what if the overall context in a causal scenario is changed in a way that renders the bogus preventer comparably abnormal? As it turns out, it is not difficult to invent such cases:

Example 3 The Assassin of the previous example has switched careers and is now a bartender. Victim walks in the bar and orders a mojito. For some unknown reason, Assassin refrains from putting in alcohol. Bodyguard puts in special alcohol neutralizer nevertheless. Victim drinks the mojito and remains sober.

³ By this I mean that the implausibility need not arise from one's theoretical commitments, such as the correspondence hypothesis about judgments of actual causation.

The variables could be:

$$\begin{aligned}
 A &= \begin{cases} 1 & \text{if Assassin puts in alcohol} \\ 0 & \text{otherwise} \end{cases} \\
 B &= \begin{cases} 1 & \text{if Bodyguard puts in special alcohol neutralizer} \\ 0 & \text{otherwise} \end{cases} \\
 VS &= \begin{cases} 1 & \text{if Victim remains sober} \\ 0 & \text{if she doesnt} \end{cases}
 \end{aligned}$$

The equation describing the structure is:

$$VS = \max((1 - A), B)$$

The witness $s_{\vec{w},\vec{x},\vec{u}}$ for $B = 1$ being an actual cause of $VS = 1$ is $(A = 1, B = 0, VS = 0)$ and the actual small world $s_{\vec{u}}$ is $(A = 0, B = 1, VS = 1)$. Intuitively, it is now the case that $s_{\vec{w},\vec{x},\vec{u}} \succeq s_{\vec{u}}$ given that variables A and B take more normal values in $s_{\vec{w},\vec{x},\vec{u}}$.⁴ Thus, worlds $s_{\vec{w},\vec{x},\vec{u}}$ and $s_{\vec{u}}$ count as comparable and the witness $s_{\vec{w},\vec{x},\vec{u}}$ is more normal than the actual world $s_{\vec{u}}$. Therefore, $B = 1$ counts as an actual cause of $VS = 1$ under Halpern and Hitchcock’s account. This is so even though it seems clear that Bodyguard’s putting in alcohol neutralizer doesn’t play any role in Victim’s remaining sober if there is no alcohol in Victim’s drink in the first place. Similar scenarios are very easy to invent by making sure that the value of the effect variable is normal (I’ll discuss one such example later). This suggests that the incomparability strategy is not the correct remedy to Example 2 either, but simply happens to deliver the correct result in that case, perhaps accidentally.

If it is true that Halpern and Hitchcock’s formal model has trouble with a particular class of causal scenarios, namely those cases in which the effect variable takes an intuitively normal value, then this points to the possibility that one of the underlying assumptions is wrong. One of the immediately suspect parts of Halpern and Hitchcock’s theory is the idea that actual causes are always abnormal. Let us now examine this issue in detail.

4 The Psychology of Normal Effects

In the previous section, we saw that Halpern and Hitchcock’s account of actual causation struggles with certain types of scenarios. In this section, I argue that the reason is that Halpern and Hitchcock are committed to the wrong analysis of actual causation, namely the idea that actual causes are always *abnormal*. I will examine the details of this commitment and show how it can lead to a tension with certain psychological theories about causal judgments. I will then present a particular psychological approach with which Halpern and Hitchcock’s account is in tension,

⁴ For the purposes of this example, we can work with Halpern and Hitchcock’s (2013) assumption that the normal value of the effect variable VS is the one determined by A and B according to the relevant structural equation. In that case, the value of VS is normal in both worlds under comparison and can be therefore ignored.

namely the *correspondence* hypothesis about causal judgments. I will argue that the correspondence hypothesis is better supported by the empirical evidence and that formal models of actual causation should be rendered compatible with it. Finally, in order to address some otherwise puzzling examples, I will discuss the way in which causes can influence the normality of their effects in causal scenarios.

4.1 Philosophers' Commitment to Abnormal Actual Causes

The idea that ordinary people tend to judge abnormal events as actual causes is not new in philosophy (Hart and Honoré 1985). However, problem cases such as the ones discussed in the first half of this paper have led to various explicit and increasingly formal attempts to build normality considerations into accounts of actual causation (cf. Menzies 2007; Hall 2007; Hitchcock 2007). The account put forward by Halpern and Hitchcock (2014) can be seen as the culmination of these efforts.

To understand the role of abnormality considerations in these recent accounts, it is helpful to think about the influential paper by Hitchcock and Knobe (2009). In this paper, the authors not only argue that abnormal events are generally judged as actual causes but also provide a set of justifications for why this should be so. For example, Hitchcock and Knobe argue that in mentally 'undoing' some effect that has actually occurred, it makes sense to keep in one's imagination the preceding normal events fixed and vary the preceding abnormal events to see if there would be a change in the effect. The reason it makes sense to do this and not the opposite, namely to vary the normal events while keeping the abnormal events fixed, is that the abnormal events are less likely to occur in the future. Therefore, the situation where abnormal events are varied and normal events are kept fixed is more generalizable out of these two alternatives (Hitchcock and Knobe 2009, p. 607). The arguments put forward by Hitchcock and Knobe sparked a fascinating debate on the role (if any) of different types of normality and typicality in judgments of actual causation (Samland and Waldmann 2014; Livengood and Rose 2015; Livengood et al. 2016; Sytsma et al. 2012). Later I will discuss some of the results presented in these debates.

It will be helpful to clarify the specific way in which the accounts of actual causation discussed above are committed to abnormal actual causes. Consider the following three questions:

- (a) What are the psychological mechanisms explaining why events (or other causal relata) are judged as normal or abnormal?
- (b) What are the psychological mechanisms explaining why ordinary people tend to pick out abnormal events (or other causal relata) as actual causes?
- (c) How are normal and abnormal causes related with normal and abnormal effects in people's causal judgments?

A formal framework for causal judgments, such as the one put forward by Halpern and Hitchcock, could plausibly make commitments with respect to all three of the above questions, or with respect to none of them. In practice, the philosophical

accounts discussed above have largely left questions (a) and (b) to psychologist. For example, Hitchcock and Knobe say that the precise details of the reasoning processes underlying judgments of actual causation do not matter for their purposes (Hitchcock and Knobe 2009, p. 599). Halpern and Hitchcock (2014) seem similarly flexible about questions (a) and (b). Neutrality with respect to (a) and (b), however, does not imply neutrality with respect to (c), as can be seen from our discussion on Hitchcock and Knobe (2009) above.

Similarly, as the discussion in the first part of this paper suggests, Halpern and Hitchcock (2014) build it into their framework that actual causes are abnormal events (values of variables), even though Halpern and Hitchcock otherwise remain somewhat ecumenical about the underlying psychology. To see Halpern and Hitchcock's commitment with respect to (c) very clearly, consider that in order for a counterfactual relationship between variables $X = x$ and $Y = y$ to count as causal under Halpern and Hitchcock's definition of actual causation, the witness world for that relationship, call it $w_{x' \rightarrow y'}$, needs to be at least as normal than the actual world where the relationship holds. Now imagine that X is a binary variable with just two values, x and x' , where x is the normal and actual value of X . In this case, X takes a less normal value in $w_{x' \rightarrow y'}$. Therefore, if $X = x$ is to count as the actual cause of $Y = y$, then there must be some *other* variables in the causal model which take abnormal values in the actual world and normal values in $w_{x' \rightarrow y'}$. One of the consequences of this is that $X = x$ cannot count as the cause of $Y = y$ in the simple causal model consisting of just these two variables. Another consequence is that, if all the cause variables in a model are like X (i.e. binary and normal in the actual world), then none of those variables can count as actual causes, since the witnesses for the relevant counterfactuals cannot be at least as normal as the actual world. There are plausibly other similar consequences which are more laborious to demonstrate.⁵

I said above that Halpern and Hitchcock are 'somewhat ecumenical' about the underlying psychology. There is a reason for the word 'somewhat': as I will now demonstrate, Halpern and Hitchcock's commitments with respect to question (c) lead into a conflict between the formal framework they put forward and a particular empirically plausible psychological theory about judgments of actual causation.

⁵ As an example, imagine that, in addition to X , there *are* some variables in the relevant causal model that take abnormal values in the actual world and normal values in the witness. Call any such variable Z_i . I will now argue that, in cases like these, there is always some Z_i that will count as an actual cause of $Y = y$ according to conditions AC1–AC3 and with the witness $w_{x' \rightarrow y'}$. Conditions AC1 and AC2(a) are satisfied by assumption, since we require both that Z_i takes an abnormal value z_i in the actual world where $Y = y$ holds, and that it takes a normal value $z'_i \neq z_i$ in $w_{x' \rightarrow y'}$ where $Y = y' \neq y$ holds. To see that condition AC2(b) is satisfied with X in \vec{W} and Z_i in \vec{Z} , suppose this was *not* the case. Then we would have had no reason to consider, when we wanted to uncover the counterfactual dependency between $X = x' \neq x$ and $Y = y'$, a world such as z'_i in $w_{x' \rightarrow y'}$ where Z_i takes a non-actual value. In other words, the fact that Z_i takes value z'_i in the witness implies that $Y = y$ holds when $Z_i = z_i$ even when $X = x'$, just as required by AC2(b). Condition AC3 is similarly satisfied for Z_i under the assumption that $Z_i = z'_i$ must hold in $w_{x' \rightarrow y'}$. This means that a normal cause such as $X = x$ can never counts as the *best* actual cause for $Y = y$, although it can in some cases count as an equal actual cause together with some variable with an abnormal actual value.

4.2 The Correspondence Hypothesis About Actual Causation

Halpern and Hitchcock write that their account can be seen ‘as a formalization of Kahneman and Miller’s (1986) observation that we tend to consider only possibilities that result from altering atypical features of a world to make them more typical, rather than vice versa’ (Halpern and Hitchcock 2014, pp. 435–436). This is the commitment to abnormal actual causes discussed above. But a closer look at *Kahneman and Miller’s* theory makes it very clear that, for them, those ‘atypical features’ include *effects* as well as causes:

Causal questions about particular events are generally raised only when these events are abnormal. [...] A [causal] why question indicates that a particular event is surprising and requests the explanation of an *effect*, defined as a contrast between an observation and a more normal alternative. (Kahneman and Miller 1986, p. 148, original italics.)

So while Kahneman and Miller’s research seems to provide support for the idea that people usually treat abnormal events as actual causes, it also suggests that the kinds of effects on which people usually focus are abnormal events. As we recall from the earlier discussion, this ‘abnormality bias’ was identified and criticized by psychologists in the 80s (cf. Einhorn and Hogarth 1986; Gavanski and Wells 1989). What these researchers proposed instead was the *correspondence hypothesis* about causal judgments, which implied not only that abnormal effects are judged to have abnormal causes but also *the opposite*: that normal effects are judged to have normal causes. Now the reason why this hypothesis is interesting from the point of view of our discussion is that, in the problematic prevention cases discussed above, the *effect variable seems to take a value that is normal*. Surviving, for example, can be thought of as a normal state for Victim to be in, and therefore it is not clear whether its actual causes have to be something abnormal, any more than the actual causes of my getting to work at the usual time today need to be something abnormal.

As I mentioned earlier, the correspondence hypothesis about causal judgments was also empirically tested. Gavanski and Wells (1989) asked respondents to evaluate a set of causal scenarios in which the value of the effect variable was either normal or abnormal. Below is an example of one such scenario (Gavanski and Wells 1989, p. 319), with the variable features of the story bracketed and my numbering scheme:

1. Andrea is a first year Sciences student in university. She is (a very poor) student. She obtained a (2. 1) grade point average in her first term.
2. Andrea had an important exam coming up on Friday in her psychology course. She spent about her usual amount of time studying the class notes and assigned readings. (Although she rarely felt this was necessary) Andrea went to the library to pick up some extra reference material for the exam, which she studied carefully.
3. On Wednesday, (Andrea picked up a copy of the previous year’s exam from a friend who had taken the same psychology course last year. This was something

- she commonly did for important exams.) Andrea studied the aid exam carefully to get an idea of the kinds of questions that might be on her exam.
4. (Andrea's mother was ill on Thursday and unable to work around the house. Because of this, Andrea spent Thursday evening doing household chores—something she rarely did.)
 5. On Friday morning, Andrea was very nervous, (as she always was just before important exams). She noticed that her hands were shaking and hoped that her anxiety would not interfere with her concentration.
 6. Andrea wrote the exam later that day and (failed).

Here, the first paragraph sets an expectation about the likely outcome of the process (described in paragraph 6), which is that Andrea will fare poorly. Paragraphs 2 and 4 describe certain abnormal events that happen in the days leading to the exam. Paragraphs 3 and 5, by contrast, describe certain normal events that take place in those days. The normality of the bracketed sections were varied so that for some respondents the events in paragraph 2 and 4 were normal and those in paragraphs 3 and 5 were abnormal. Finally, Gavanski and Wells varied the normality of the outcome of the process by changing whether Andrea was a good student or what the outcome of the exam was. 229 subjects were then asked to list six things that could have changed the outcome of the scenario.

What Gavanski and Wells found was that, when the outcome of the process was normal (e.g. an excellent student passing an exam), subjects tended to vary the normal aspects of the story towards abnormality in order to 'undo' the outcome. But when the outcome was abnormal (e.g. an excellent student failing an exam), subjects tended to vary the abnormal aspects of the story towards normality in order to undo the outcome.⁶ Here is how Gavanski and Wells (1989, p. 321) summarize their results:

This research indicates that mental simulation is governed by a correspondence between the normality of outcomes and prior events. Previous research, which has used only exceptional outcomes, has been interpreted as evidence that people make mutations primarily in the direction of normality. The current research found this effect for exceptional outcomes but found the opposite effect for outcomes that were normal.

More recently, the *normality*-correspondence hypothesis has been generalized even further, with some studies suggesting that there are other important features in addition to normality, such as magnitude, that people seek to mentally 'match' between causes and effects (Bouts et al. 1992; Sim and Morris 1998). These discoveries suggest that our future models of actual causation may ultimately need

⁶ Gavanski and Wells also found that people were generally more likely to vary those aspects of the scenario that were abnormal. However, they varied those aspects in both directions. When the outcome was normal, subjects tended to vary abnormal events towards normality; but when the outcome was abnormal, subjects tended to vary abnormal events towards *further* abnormality. As Gavanski and Wells correctly point out, the likely explanation for this is that it is easy to imagine more and less abnormal alternatives to abnormal events, while it is harder to imagine more normal alternatives to normal events (Gavanski and Wells 1989, p. 322).

to incorporate a greater number of considerations than just normality-correspondence. Although such considerations are beyond the scope of this paper, they seem like a fascinating opportunity for future research.

The correspondence hypothesis is also compatible with some of the recent results from experimental philosophy. Recall our earlier discussion about the account of actual causation put forward by Hitchcock and Knobe (2009), see also Knobe and Fraser (2008). Several philosophers have presented studies arguing against Knobe and Hitchcock's abnormality-based account (cf. Livengood and Rose 2015; Livengood et al. 2016; Sytsma et al. 2012; Samland and Waldmann 2014). While the criticism presented in these papers were different from the worries raised here, the empirical evidence these experimental philosophers used to support their arguments is very interesting from our point of view. For what authors such as Sytsma et al. (2012) found was that, in certain experimental scenarios, people's judgments were *not* sensitive to the 'typicality' (or in our terminology 'normality') of the putative actual causes in the way postulated by Knobe and Hitchcock and others. Sytsma et al. (2012) argued that this is because there is a distinction between between population-level and agent-level typicality such that causal attributions are in fact sensitive to agent-level typicality (rather than atypicality). This is interesting because it is plausibly another dimension along which causal judgments vary. But another relevant observation from our point of view is that, in the vignettes presented by Sytsma et al. (2012, Appendix), the normality or abnormality of the *effect* was not specified. Because respondents' assumptions about the normality or abnormality of the effect was allowed to vary, the results are therefore compatible with the correspondence hypothesis. In fact, the results could be partly explained by the hypothesis.

4.3 The Tension Between Halpern and Hitchcock's Account and the Correspondence Hypothesis

As we recall from Sects. 2 and 3, Halpern and Hitchcock supplement their account of actual causation with normality considerations in order to deal with certain example scenarios such as bogus prevention, where definition of actual causation they put forward would otherwise fail to deliver the intuitively correct result. Specifically, they require that the witness worlds for the relevant counterfactual causal relations must be at least as normal the actual world is. This feature of their definition of actual causation then helps Halpern and Hitchcock to capture seemingly ordinary intuitions about a number of causal scenarios. As suggested above, Halpern and Hitchcock are able to make the above commitment without simultaneously adopting any particular psychological theory about normality or causal judgments, although of course they cite Kahneman and Miller's theory as their inspiration. However, as may now be obvious, their commitment to abnormal actual causes is nevertheless not compatible with certain psychological theories about causal judgments, such as the correspondence hypothesis discussed above. To see this clearly, consider the following shortened version of the story of Andrea presented above:

Andrea is a very good student. As is usual, she spends the night studying the course materials. In the hours leading to the exam, she is calm and focused. Typically for her, she sits the exam and passes with excellent marks.

Now we might want to say that the actual causes for Andrea's passing with excellent marks ($PASS = 1$) is her preparing in the previous evening ($PREPARE = 1$) and her being calm and focused ($FOCUS = 1$) in the morning of the exam. Suppose the causal structure is

$$PASS = \min(PREPARE, FOCUS)$$

with all variables having two values, 1 for 'occurs' and 0 for 'doesn't occur'. Then it seems like preparing and focusing cannot be the actual causes of passing under Halpern and Hitchcock's account because the contingency that we would need to consider to establish the counterfactual would require setting all of these variables into their abnormal values.

Now, perhaps Halpern and Hitchcock could appeal to some unorthodox theory of 'normality' to render the intuitively 'normal' causes in the above example as 'abnormal'.⁷ After all, Halpern and Hitchcock do not commit themselves to any particular analysis of normality. While this type of manoeuvre would be *logically* possible and would deal with the problem of 'normal' causes (by demonstrating that they are in fact 'abnormal'), it is unclear why any psychologist would want to develop an account of normality which contradicts both ordinary use and intuitive judgment. It is also unlikely that philosophers and scientists working on actual causation would want to endorse this theory since one of the criteria for a successful account of actual causation is that it captures ordinary intuitions about causal scenarios. This is presumably also why Halpern and Hitchcock spend the majority of their paper demonstrating how their definition of actual causation captures such intuitions.

To summarize, then, my argument is that Halpern and Hitchcock's commitment to abnormal actual causes isn't compatible with the empirical evidence—barring the adoption of some very unorthodox theory about 'normality' which doesn't exist and is unlikely to be developed. The same is true for the several other abnormality-based philosophical accounts of actual causation mentioned above.

4.4 The Correspondence Hypothesis and Formal Modeling

I have above presented several critical remarks about Halpern and Hitchcock's account. This does not imply that I regard their approach in some way fundamentally flawed. In fact, it seems plausible to me that the structural equations framework with *some* kind of normality metric could incorporate the empirical results that I have discussed, as well as the correspondence hypothesis about judgments of actual causation that seems to underlie those results. However, this requires giving up the commitment to the idea that actual causes are abnormal

⁷ This type of possibility is suggested by an anonymous referee.

events or values of variables, which is an assumption in Halpern and Hitchcock's account and in a number of other philosophical theories.

How significant is this requirement? For some accounts of actual causation, such as the one put forward by Hitchcock and Knobe (2009), the requirement is plausibly fairly significant, since the authors provide various reasons for why it should be the case that people focus on abnormal events as actual causes. For Halpern and Hitchcock (2014), on the other hand, the question is plausibly that of whether there is a natural way of modifying the requirement that the witness worlds revealing the counterfactual dependencies between variables need to be at least as normal as the actual world, such that the requirement is either dropped or changed in some other way when the effect variable is normal. Investigating this possibility is a task for future research.

In the following final section of the paper, I will briefly discuss the factors influencing the perceived normality of an effect. This is clearly an important issue from the point of view of the correspondence hypothesis since it postulates that the normality of actual causes depends on the normality of the actual effect. We will also deal with some example scenarios which might otherwise seem puzzling in the light of the correspondence hypothesis.

4.5 Causes, Effects and Normality Dynamics

Recall the experiment that Gavanski and Wells (1989) conducted about judgments of actual causation when the effect is normal. In the paragraph they presented to their respondents, the normality or abnormality of the outcome (good or bad exam success) was determined by the context set in the first paragraph (good or bad student). The fact that such simple reframing serves to alter respondents' judgments suggests that the perceived normality of an effect is rather sensitive to the context. In fact, this observation had already been made a few years earlier by Wells and others (Wells et al. 1987). Based on their experimental work, Wells et al. (1987, p. 429) hypothesized that the perceived normality of an outcome in a causal scenario depends on the prior events in that scenario. For example, in the simplest case, *given* that someone pulled the trigger, it was normal for the gun to fire; whereas *given* that no one pulled the trigger, the gun's firing was perceived as abnormal. Furthermore, Wells and his colleagues observed in an experiment that the *first* event in a causal chain seemed to constrain the availability of alternatives for the subsequent events in the chain. In the light of the norm theory discussed earlier, this implies that the first event in a causal chain plays a particularly important role in influencing the perceived normality of those subsequent variables that are causally related with the event. This idea seems to explain why, when asked to 'undo' the outcome of a causal chain leading to an abnormal outcome, subjects tended to alter the first event in the causal chain, rather than the subsequent ones, to change the outcome.

The question now arises: how is it *ever* possible for an effect to be abnormal? Surely, in most situations, the occurrence of an effect is normal given that its causes have occurred. Indeed, the apparent fact that causes typically reduce the 'surprisingness' of their effects has been used to link causality with explanation (Hausman 1993). Does this mean that *all* effects are normal? The answer to this

question is nuanced. It is true that, given the occurrence of its causes, the occurrence of an effect is usually in some sense normal.⁸ This is plausibly what links causality with explanation. At the same time, effects can be abnormal in the light of many other contextual factors, causal or otherwise. Gunshots, for example, can count as abnormal relative to statistical, social and many other norms, even if they count as normal relative to the fact that someone pulled the trigger. A train's arriving at the usual time, by contrast, tends to count as normal relative to its causes *and* whatever other factors are relevant for the normality judgment, although of course there are situations in which that might not be so.⁹

The apparent fact that causes seem to constrain the available alternatives for (and hence the normality of) their effects in this way seems to result in an interesting kind of dynamic in some prevention scenarios. Take the simplest case in which there is one cause variable, one effect variable and one variable preventing the influence of the cause on the effect. In accordance with the observations made by Wells et al. (1987), the value of the cause variable—even in the presence of a preventer—can *still* constrain the available alternatives for (that is: the normality of) the value of the effect variable. *Given* that the cause occurred, we may judge that it would be normal for the effect to occur. If the effect then fails to occur because of the presence of the preventer, this non-occurrence can have highly available alternatives (that is: can be judged abnormal).

Let us now consider different types of prevention scenarios to see whether the above observations seem to be along the right lines. We shall start with a case that seems to pose a challenge to the claim that Halpern and Hitchcock's account struggles specifically with those cases in which the effect variable takes a normal value:

Example 4 A village in Southern England gets regularly flooded in the spring, causing minor damage in some of its medieval buildings. This year, the river banks are made higher as the accidental consequence of some building works. As it happens, the spring is exceptionally dry and there are no floods. The medieval buildings remain undamaged.

As before, we can represent the scenario with the following variables:

$$A = \begin{cases} 1 & \text{if village gets flooded} \\ 0 & \text{otherwise} \end{cases}$$

$$B = \begin{cases} 1 & \text{if river banks are made higher} \\ 0 & \text{otherwise} \end{cases}$$

$$D = \begin{cases} 1 & \text{if medieval buildings are damaged} \\ 0 & \text{otherwise} \end{cases}$$

Let us suppose that the causal structure of the situation is as follows:

⁸ I am setting aside cases where, for example, the occurrence of the effect is still highly unlikely.

⁹ Incidentally, this may be why causal explanations of normal effects can sometimes feel disappointing. A causal explanation for why the train arrived at its usual time may fail to reduce any element of surprisingness involved, simply because that element may have been non-existent to begin with.

$$D = A \times (1 - B)$$

First we can easily confirm that the above example is a problem for Halpern and Hitchcock's account, for all the familiar reasons. Under their model, $B = 1$ counts as an actual cause of $D = 0$ with the witness ($A = 1, B = 0, D = 1$), which is more normal than the actual world ($A = 0, B = 1, D = 0$). $B = 1$'s being an actual cause of $D = 0$ is, of course, intuitively incorrect. But note that we can also observe that the occurrence of minor damage to the village's medieval buildings, i.e. $D = 1$, was supposed to be *normal*. In the light of *this*, $D = 0$ appears abnormal. In the absence of the above considerations on how causes can constrain the normality of effects, this example might be seen as going against the argument that Halpern and Hitchcock's account struggles with cases in which the value of the effect variable is *normal*.

Here is another scenario which might appear similarly initially puzzling:

Example 5 Assassin puts a lethal dose of poison in Victim's coffee. Moments later, Bodyguard decides to put in generic antidote to Victim's coffee even though she has no reason to suspect any foul play. She is acting on a whim. Victim then drinks the coffee and survives.

We can model the scenario with the obvious variables:

$$A = \begin{cases} 1 & \text{if Assassin puts in poison} \\ 0 & \text{otherwise} \end{cases}$$

$$B = \begin{cases} 1 & \text{if Bodyguard puts in antidote} \\ 0 & \text{otherwise} \end{cases}$$

$$VS = \begin{cases} 1 & \text{if Victim survives} \\ 0 & \text{otherwise} \end{cases}$$

And the equation:

$$VS = \max((1 - A), B)$$

Why might the scenario appear puzzling? Well, here we have a case in which a value of a variable ($B = 1$) which is in most circumstances abnormal causes a value of a variable ($VS = 1$) which is in *most circumstances* normal. In the absence of further arguments, it might seem as if we would have an obvious counterexample to the correspondence hypothesis about actual causation.

However, the observations by Wells et al. (1987, p. 429) about the context-sensitivity of normality explain the above puzzles in a rather straightforward manner. In Example 4, even though the buildings' getting damaged ($D = 1$) is statistically normal, it isn't normal *given* that there is no flood ($A = 0$). Thus, as predicted by Wells et al. (1987), a prior event in a causal process constrains the available alternatives for the actual value of D , rendering $D = 0$ normal and in so doing trumping the *statistical* norm.¹⁰ Note, however, that the river banks' being

¹⁰ This observation seems intuitively correct and is compatible with experimental results. However, as far as I know, no experiment has pitted statistical norms and what we might call 'causal norms' directly against one another. This would be an interesting topic for future study.

accidentally made higher ($B = 1$) is *not* causally related with whether there is flood or not (A). Therefore, the value of A does *not* constrain the alternatives available for $B = 1$. The value of B is consequently perceived as abnormal on statistical and other grounds, resulting in a mismatch between the normalities of $B = 1$ and $D = 0$. As the result, and in contrast with the verdict delivered by Halpern and Hitchcock's theory, $B = 1$ is not judged an actual cause of $D = 0$.

Similar observations apply to Example 5. Here, Assassin's putting in poison seems to constrain the value of the outcome variable. For instance, it seems to be an important piece of information that the dose is *lethal*. This serves to further specify the way in which Assassin's putting in poison constrains the alternatives for the value of the outcome variable. Suppose the dose was *non-lethal*; it seems that our expectations about Victim's survival would be different. *Given* that Assassin puts in lethal poison ($A = 1$), it is normal for VS to take value 0. Against this backdrop, $VS = 1$ turns out abnormal. Further, we can observe that the normality of the value of B is not influenced by the value of A since Assassin's putting in poison does not, in the absence of further information, influence the availability of the alternatives for Bodyguard's putting in antidote. The normality of the value of B is consequently determined by other factors, including statistical norms. Therefore, both $B = 1$ and $VS = 0$ are abnormal values of variables in this scenario.

In these types of simple prevention scenarios with one cause, one preventer and one effect, the presence of the preventer seems to elicit an evaluation of the normality of the effect in the light of the value of the cause variable. It seems that, in the *absence* of the preventer, the normality of both the cause and the effect in these scenarios would be evaluated in relation to other kinds of factors, such as statistical norms and other features discussed in the previous section. For the sake of convenience (and for that reason *only*) I will below call these other normality-influencing factors 'external'. Thus, if one imagines Examples 3, 4 and 5 without the preventing variable, they seem like cause-effect pairs where both of the variables take abnormal values, where this abnormality is judged in relation to external factors. However, when a preventer is present, it seems that the value of the cause variable begins to constrain our judgment about the normality of the value of the effect variable. Thus, if one has the intuition that, say, in Example 4 the absence of flood ($A = 0$) causes the absence of damage ($D = 0$), it seems that this is because the relationship between the *pair* of variables (A and D) is evaluated in relation to external factors, which is when both values count as abnormal.¹¹ However, when all *three* variables (A , B and D) are evaluated together as a prevention scenario, the normality of the effect variable ($D = 0$) is strongly constrained by the value of the cause variable ($A = 0$). If people have mixed intuitions about actual causation in these types of prevention scenarios, a possible explanation is that normality and representativeness can depend on the frame of reference in this way.

Finally, it is interesting to consider a prevention scenario in which the value of the cause variable seems to constrain the value of the *preventing* variable. Fuses are

¹¹ This intuition is probably not universal. After all, there is already a preventer (higher river banks) in place. The same is true in Example 3. In the case of genuine prevention (Example 5), the intuition doesn't seem to occur at all.

devices whose purpose is to prevent damage caused by overcurrent. Thus, we might have the following type of case:

Example 6 An overcurrent flows through a wire, causing a fuse to blow. An electronic device attached to the other end of the wire remains undamaged.

Here are our variables:

$$\begin{aligned}
 OC &= \begin{cases} 1 & \text{if overcurrent flows through wire} \\ 0 & \text{otherwise} \end{cases} \\
 FB &= \begin{cases} 1 & \text{if fuse blows} \\ 0 & \text{otherwise} \end{cases} \\
 ED &= \begin{cases} 1 & \text{if electronic device is damaged} \\ 0 & \text{otherwise} \end{cases}
 \end{aligned}$$

The following equations describe the causal structure of the scenario:

$$\begin{aligned}
 OC &= U \\
 FB &= OC \\
 ED &= OC \times (1 - FB)
 \end{aligned}$$

Here, the value of OC seems to constrain the availability of alternatives for $FB = 1$: the occurrence of the overcurrent plausibly reduces the easiness with which alternatives to the fuse's blowing come to mind. Consequently, $FB = 1$ appears to be a normal value of the variable given that $OC = 1$. However, it now similarly appears that $ED = 0$ counts as normal. If it is normal for the fuse to blow, it is hard to see why it would not count as normal for there to be no damage for the electronic device attached to the other end of the wire. Once again, the intuitive actual cause seems to correspond in normality with the effect.

The general observation arising from the three examples discussed in this section is that the normality of a value of a variable in a causal model can be influenced by the values of the other variables in the model. While this in itself is not particularly surprising, it seems to result in interesting kinds of dynamics in certain prevention scenarios. In this section, I have argued that, when those dynamics are taken into account, the correspondence hypothesis about actual causation seems to explain our intuitive judgments about the three types of prevention scenarios discussed above. In my view, this further strengthens the case I have made in support of the correspondence hypothesis and against abnormality-based accounts. At the same time, it is clear that many of the hypotheses above can be tested and further developed in future work.

5 Conclusion

After presenting a psychological theory about counterfactual judgments, Kahneman and Varey (1990, p. 1108) observed that their work occupies 'a somewhat uncomfortable middle ground between psychology and philosophy'. In their view,

philosophers seek to ‘impose a consistent logical structure’ on causality, probability and counterfactual conditionals, while ‘the study of human thinking should neither assume nor impose consistency on its subject matter’ (p. 1108). The discussion in this paper has also occupied the middle ground between philosophy and psychology, although I leave it for the reader to decide whether this territory is ‘uncomfortable’. In so far as our goal is to form rigorous *and* realistic models of judgments of actual causation, it seems that this kind of interaction between the conceptual and the empirical is unavoidable. As Kahneman and Varey (1990, p. 1108) already point out, ‘persuasive philosophical arguments commonly draw on compelling examples that evoke strong shared intuitions’.

In this spirit, and as we remember, the main ‘evidence’ that Halpern and Hitchcock present for their formal account of actual causation, which I described in Sect. 2, is that it seems to deliver the intuitively correct verdict about the actual causes in a number of example scenarios. I briefly mentioned some of these examples in Sect. 3, although my main focus there was to discuss those cases that Halpern and Hitchcock’s model does *not* get right. These examples, it turned out, are ones in which the effect variable takes what intuitively seems like a *normal* value. In Sect. 4, I presented my diagnosis about what the problem is. I pointed out that Halpern and Hitchcock are committed to an *abnormality*-based approach to actual causation. I then argued in favour of an alternative view, a *correspondence*-based theory. I described a tension between Halpern and Hitchcock’s approach and the correspondence hypothesis, arguing that this tension should be relieved by giving up the commitment to the idea that actual causes are always abnormal. Finally, I dealt with some initially puzzling examples by discussing the way in which causes can constrain the alternatives available for their effects, and how this seems to influence our judgments in certain prevention scenarios.

Acknowledgements Parts of this paper were presented in the PSA 2014 conference in Chicago and in the Causal and Probabilistic Reasoning conference in Munich in 2015. I thank the participants of these conferences for stimulating discussions. I also thank David Papineau and Eleanor Knox for their many valuable insights and three anonymous reviewers of this journal for their helpful comments.

Funding Funding was provided by Koneen Säätiö.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bouts, P., Spears, R., & Van Der Pligt, J. (1992). Counterfactual processing and the correspondence between events and outcomes: Normality versus value. *European Journal of Social Psychology*, 22, 387–396.
- Danks, D. (2013). Functions and cognitive bases for the concept of actual causation. *Erkenntnis*, 78, 111–128. doi:10.1007/s10670-013-9439-2.

- Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin*, 99, 3.
- Gavanski, I., & Wells, G. L. (1989). Counterfactual processing of normal and exceptional events. *Journal of Experimental Social Psychology*, 25, 314–325.
- Glymour, C., Danks, D., Glymour, B., Eberhardt, F., Ramsey, J., Scheines, R., et al. (2010). Actual causation: A stone soup essay. *Synthese*, 175, 169–192. doi:10.1007/s11229-009-9497-9.
- Hall, N. (2007). Structural equations and causation. *Philosophical Studies*, 132, 109–136. doi:10.1007/s11098-006-9057-9.
- Halpern, J. Y. (2013). Appropriate causal models and stability of causation. Technical report.
- Halpern, J. Y. (2016). *Actual causality*. MIT Press. URL <http://www.jstor.org/stable/j.ctt1f5g5p9>.
- Halpern, J. Y., & Hitchcock, C. (2011). Actual causation and the art of modeling. arXiv:1106.2652.
- Halpern, J. Y., & Hitchcock, C. (2013). Compact representations of extended causal models. *Cognitive Science* 37:6. arXiv:1309.1227 [cs]. 2013 (pp 986–1010).
- Halpern, J. Y., & Hitchcock, C. (2014). Graded causation and defaults. *The British Journal for the Philosophy of Science*, 66, 413–457.
- Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science*, 56, 843–887.
- Hart, H. L. A., & Honoré, T. (1985). *Causation in the law*. Oxford: Clarendon Press.
- Hausman, D. M. (1993). Linking causal and explanatory asymmetry. *Philosophy of Science*, 60, 435–451.
- Hitchcock, C. (2007). Prevention, preemption, and the principle of sufficient reason. *The Philosophical Review*, 495–532. <http://www.jstor.org/stable/20446988>.
- Hitchcock, C., & Knobe, J. (2009). Cause and norm. *The Journal of Philosophy*, 106, 587–612.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93, 136.
- Kahneman, D., & Tversky, A. (1981). The simulation heuristic. Technical report DTIC document.
- Kahneman, D., & Varey, C. A. (1990). Propensities and counterfactuals: The loser that almost won. *Journal of Personality and Social Psychology*, 59, 1101.
- Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. *Moral Psychology*, 2, 441–448.
- Livengood, J., & Rose, D. (2015). Experimental philosophy and causal attribution. *A Companion to Experimental Philosophy*, 434–449.
- Livengood, J., Sytsma, J., & Rose, D. (2016). Following the fad: Folk attributions and theories of actual causation. *Review of Philosophy and Psychology*. doi:10.1007/s13164-016-0316-1.
- Menzies, P. (2007). Causation in context. In: R. Corry & H. Price (Eds.), *Causation, physics, and the constitution of reality* (pp. 191–223). Oxford: Oxford University Press.
- Mill, J. S. (1856). *A system of logic, ratiocinative and inductive, being a connected view of the principles, and the methods of scientific investigation* (Vol. 2). Chicago: JW Parker.
- Prentice, R. A., & Koehler, J. J. (2002). A normality bias in legal decision making. *Cornell Law Review*, 88, 583.
- Samland, J., & Waldmann, M. R. (2014). Do social norms influence causal inferences. In *Proceedings of the 36th annual conference of the cognitive science society* (pp. 1359–1364).
- Sim, D. L., & Morris, M. W. (1998). Representativeness and counterfactual thinking: The principle that antecedent and outcome correspond in magnitude. *Personality and Social Psychology Bulletin*, 24, 595–609.
- Sytsma, J., Livengood, J., & Rose, D. (2012). Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, 814–820.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207–232.
- Wells, G. L., Taylor, B. R., & Turtle, J. W. (1987). The undoing of scenarios. *Journal of Personality and Social Psychology*, 53, 421.