Making a Difference

Making a Difference

Essays on the Philosophy of Causation

EDITED BY
Helen Beebee,
Christopher Hitchcock,
and Huw Price

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2017

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2016962739

ISBN 978–0–19–874691–1

Printed in Great Britain by
Clays Ltd, St Ives plc

In memory of Peter Menzies

10

Cause without Default

Thomas Blanchard and Jonathan Schaffer

> [A] cause is an intervention, analogous to a human action, that brings about changes in the normal course of events.
>
> (Menzies 2011: 356)

Must causal models distinguish *default* from *deviant* events? Much recent work on actual causation is conducted within the structural equations framework (Spirtes et al. 1993; Pearl 2000), via the notion of a causal model. In standard causal models one sets up a system of variables, allots values to these variables, and connects these variables via structural equations. Menzies (2004, 2007), Hitchcock (2007), Hall (2007), and Halpern (2008) have all argued, however, that standard causal models must be supplemented with a distinction between default (or normal, or expected) and deviant (or abnormal, or surprising) events. We aim to critically evaluate this proposal.

We grant that the notions of 'default' and 'deviant' influence causal judgement, but we claim that this influence is best understood as arising through a general cognitive bias concerning the *availability* of alternatives. (Alternatives to deviant events are more likely to leap to mind.) So we 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 availability effects on probability judgements, to try to incorporate the notions of default and deviant into the probability calculus itself).

We also claim that some key arguments for default-relativity rely on non-apt models. So we think that a second thing to be learned from these arguments is that more attention is needed concerning what counts as an apt causal model in the first place.

Overview: In §1 we introduce the structural equations framework and the notion of a causal model, discuss its connection to actual causation, and ask what makes a given model apt. In §2 we review the main case for incorporating default-relativity into causal models. Default-relativity is said to provide a conservative and psychologically plausible extension of standard causal modelling, in ways that solve multiple

problems. Finally in §3 we argue for excluding default-relativity from causal models. We think that default-relativity brings in complicating and under-constrained unclarities, while failing to be psychologically plausible and failing to solve the very problems it is said to solve. Overall we conclude that default-relativity belongs to the background biases of general cognitive performance, not to the specific facts of actual causation.

1 Background: Apt Causal Models for Actual Causation

We live in exciting times. By 'we' I mean philosophers studying the nature of causation. The past decade or so has witnessed a flurry of philosophical activity aimed at cracking this nut, and, surprisingly, real progress has been made. (Hitchcock 2001: 273)

1.1 Causal Models

Much recent work on actual causation is conducted within the structural equations framework (Spirtes et al. 1993; Pearl 2000), via the notion of a causal model. In standard causal models one sets up a system of variables, allots values to these variables, and links these variables via structural equations. It may be helpful to begin with a brief summary of this standard technology. (For simplicity we focus only on the deterministic case, though the technology can be fairly smoothly extended to the indeterministic case.)

Following Halpern (2000), it is helpful to distinguish three layers of structure involved in causal models. First, one introduces *the signature*, which roughly speaking describes *the situation under study*. More formally, the signature is a triple $\boldsymbol{S}=<\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{R}>$, where $\boldsymbol{U}$ is a finite set of exogenous variables modelling *the initial conditions of the system*, $\boldsymbol{V}$ is a finite set of endogenous variables modelling *the subsequent conditions of the system*, and $\boldsymbol{R}$ is a function mapping every variable $V\in\boldsymbol{U}\cup\boldsymbol{V}$ to an at-least-two-membered set of allotted values modelling the *contrast space for the conditions of the system*. Graphically these are the nodes of our system, divided into root and non-root nodes (but not yet linked by any edges), each decorated with a name and a set of multiple 'possible' values. For instance, to model a rock being thrown through a window, one might opt to work with the very simple signature $\boldsymbol{S_1}=<\boldsymbol{U_1}=\{\textit{Throw}\}, \boldsymbol{V_1}=\{\textit{Shatter}\}, \boldsymbol{R_1}>$, where $\boldsymbol{R_1}$ maps *Throw* to {0, 1} (contrasting the rock's being left unthrown with its being thrown) and maps *Shatter* to {0, 1} (contrasting the window's remaining intact with its being shattered).

On top of the signature one then introduces *the linkage*, which roughly speaking adds in *the dynamics of the system*. The linkage is a pair $\boldsymbol{L}=<\boldsymbol{S}, \boldsymbol{E}>$ where $\boldsymbol{S}$ is a signature as just characterized, and $\boldsymbol{E}$ is a set of modifiable structural equations characterizing, for every endogenous variable $V\in\boldsymbol{V}$, a function outputting a value v to V on the basis of values allotted to certain other variables, which thereby count as

the *parents* of V.[1] Each equation in E corresponds to a series of counterfactuals of the form: 'if the parents of V had taken these values, V would have taken that value'. E is also subject to the global constraint that the parenthood relations it induces never form loops. Graphically, the equations provide the directed edges between the nodes provided by the signature, under a global acyclicity constraint. In the case of the rock being thrown through the window with the signature $\boldsymbol{S_1}$ just described, a natural dynamics is $\boldsymbol{L_1}$=<$\boldsymbol{S_1}$, $\boldsymbol{E_1}$>, where $\boldsymbol{E_1}$ is {*Shatter*←*Throw*} (outputting a 0 for *Shatter* given a 0 for *Throw*, and a 1 for *Shatter* given a 1 for *Throw*).[2]

Finally, on top of the dynamics one then adds *the assignment*, which effectively says *what actually happened*. Given our focus on the deterministic case the assignment is a pair $\boldsymbol{M}$=<$\boldsymbol{L}$, $\boldsymbol{A}$> where $\boldsymbol{L}$ is the linkage as just characterized, and $\boldsymbol{A}$ is a function assigning values to every exogenous variable $V \in \boldsymbol{U}$. In the deterministic case one only needs to set the initial conditions. Graphically, the assignment function adds a further decoration to the root nodes, highlighting a unique 'actual' value. So in the case of the rock being thrown through the window one just adds $\boldsymbol{M_1}$=<$\boldsymbol{L_1}$, $\boldsymbol{A_1}$>, where $\boldsymbol{A_1}$ is the (smallest) function mapping *Throw* to 1.

So far we have built up a very simple causal model:

One Rock

$\boldsymbol{S_1}$ = <{*Throw*}, {*Shatter*}, $\boldsymbol{R_1}$>, where $\boldsymbol{R_1}$ maps both *Throw* and *Shatter* to {0, 1}
$\boldsymbol{L_1}$ = <$\boldsymbol{S_1}$, {*Shatter*←*Throw*}>
$\boldsymbol{M_1}$ = <$\boldsymbol{L_1}$, {*Throw*=1}>

Associated with every causal model is a directed acylic graph which partly conveys the causal information.[3] Suppressing all decoration save for the names on the nodes, here is the graph associated with *One Rock*:

We pause to build out one further illustrative example, which recurs in the discussion below. This is a representative case of (symmetric) overdetermination, involving two rocks being thrown through a window at the same time, each of which is individually sufficient to shatter the window:

[1] There is an assumption of discreteness here, enforced by the earlier requirement that $\boldsymbol{U}$ and $\boldsymbol{V}$ be finite. If one had a dense causal series, no variable in the series would have a direct parent at all.

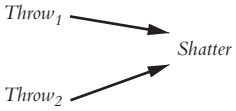
[2] Notational convention: We are using 'Φ←Ψs' to notate the idea of the value of one variable (schematically:'Φ') being determined by the values of some plurality of parent variables (schematically:'Ψs'). One sometimes sees '=' used instead, followed by a caveat that the determination in question is not really the symmetric relation of identity.

[3] Graphs associate many-one with models. Each model uniquely induces a graph. But many distinct models uniquely induce the same graph. All models with the same cardinality of variables and structure of parenthood relations induce the same graph. These graphs are thus useful but impoverished representations.

Two Rocks

$\boldsymbol{S_2}$ = <{*Throw*$_1$, *Throw*$_2$}, {*Shatter*}, $\boldsymbol{R_2}$>, where $\boldsymbol{R_2}$ maps all variables to {0, 1}
$\boldsymbol{L_2}$ = <$\boldsymbol{S_2}$, {*Shatter*←max(*Throw*$_1$, *Throw*$_2$)}> (*Shatter* gets set to 1 iff either *Throw*$_1$ or *Throw*$_2$ is at 1)
$\boldsymbol{M_2}$ = <$\boldsymbol{L_2}$, {*Throw*$_1$=1, *Throw*$_2$=1}>

The associated graph is:



1.2 Actual Causation

So far we have offered a brief summary of the notion of a causal model within the structural equations framework. We haven't yet said *anything* about what causes what. More precisely, we haven't yet said anything about any of the many notions of causation, including the relation of *actual* (or *token*, or *singular*) *causation*, which is supposed to relate one token event *c* to another token event *e* just in case *c* was in fact causally responsible for bringing about *e*. Rather we have sketched a (fruitful and elegant) framework in which various accounts of various notions of causation may be phrased.[4]

From the perspective of actual causation, the main advantage of causal models is that they permit a precise evaluation of counterfactuals whose antecedents and consequents specify situations corresponding to values of the model's variables. To evaluate such counterfactuals in a given model $\boldsymbol{M}$, one considers a modified counterpart $\boldsymbol{M}^\star$ that stipulates the new values of the variables as per the antecedent. More precisely, one may consider a counterfactual of the following schematic form, assessed in a given assigned causal model $\boldsymbol{M}$:

$$\text{If} \quad \Phi_1 = \phi_1 \quad \text{and} \quad \Phi_2 = \phi_2 \quad \text{and} \quad \Phi_3 = \phi_3 \ldots \quad \text{then} \quad \Psi_1 = \psi_1 \quad \text{and} \quad \Psi_2 = \psi_2 \quad \text{and} \quad \Psi_3 = \psi_3 \ldots$$

To assess whether this counterfactual is true in $\boldsymbol{M}$, first modify $\boldsymbol{M}$ into $\boldsymbol{M}^\star$ via the following recipe (while doing nothing further):

1. *Cut any incoming links*: For all variables Φ_j in the antecedent such that $\Phi_j \in \boldsymbol{V}$, (i) delete Φ_j from $\boldsymbol{V}$ to obtain $\boldsymbol{V}^\star$, (ii) insert Φ_j into $\boldsymbol{U}$ to obtain $\boldsymbol{U}^\star$, and (iii) delete the equation in $\boldsymbol{E}$ with Φ_j on the left to obtain $\boldsymbol{E}^\star$.

[4] Perhaps the main selling point of this framework is the development of 'discovery algorithms' that allow for causal structure to be inferred from correlational data (something which statisticians had once widely decried as impossible). The power and precision of this framework is, in our opinion, unrivalled. Not for nothing is virtually all recent work on actual causation couched in its terms. An account couched in other terms—without a development of the rival framework to comparable levels of sophistication—becomes hard to take seriously.

2. *Reassign the stipulated values*: For all variables Φ_j in the antecedent (all of which are now in $\boldsymbol{U}^*$), modify the assignment $\boldsymbol{A}$ into $\boldsymbol{A}^*$ by assigning Φ_j to the value ϕ_j specified in the antecedent.

The counterfactual is true in $\boldsymbol{M}$ if and only if the consequent ($\Psi_1=\psi_1$ and $\Psi_2=\psi_2$ and $\Psi_3=\psi_3$...) holds in $\boldsymbol{M}^*$. Effectively one has modified the model in order to surgically 'intervene' on the variables in the antecedent, by first converting them into initial conditions and then hand-setting their values.

By permitting such precise evaluation of counterfactuals, causal models permit the precise implementation of counterfactual theories of actual causation as developed previously by Lewis (1986a; cf. Menzies 1989, *inter alia*), including theories that offer precise solutions to many (and some say all) long-standing problems with overdetermination and pre-emption cases. This is an active and ongoing research programme. There is as of yet no consensus on how best to understand actual causation within the causal models framework. Indeed this is part of why there is space to argue that standard causal models should be supplemented with a default function, in order to allow for an understanding of actual causation tied into the default/deviant distinction.[5]

But for the sake of illustration, it may be useful to consider an account presented in Hitchcock (2001: 290), as it is elegant and handles many of the cases under discussion naturally. To begin with, say that there is a *directed path* from variables V_1 to V_n in model $\boldsymbol{M}$ if and only if there is a sequence of variables $<V_1, V_2,..., V_n>$ such that every variable V_j (for $1\leq j<n$) is a parent of its successor V_{j+1}. Directed paths are parenthood chains. Hitchcock's account can then be formulated as follows:

(*Hitchcock*) $X=x$ is an actual cause of $Y=y$ if and only if there is an apt causal model $\boldsymbol{M}$ such that:

1. The actual values of X and Y in $\boldsymbol{M}$ are x and y, respectively.
2. There is a directed path $\boldsymbol{P}$ from X to Y and a possible assignment of values $\boldsymbol{z}$ to the set of variables $\boldsymbol{Z}$ off $\boldsymbol{P}$ such that the following counterfactuals are true:
   a. Had $\boldsymbol{Z}$ taken values $\boldsymbol{z}$, the variables on $\boldsymbol{P}$ (except possibly X) would still have taken their actual values.
   b. Had $\boldsymbol{Z}$ taken values $\boldsymbol{z}$ and X value x, Y would have taken value y.
   c. Had $\boldsymbol{Z}$ taken values $\boldsymbol{z}$ and X some value $x^*\neq x$, Y would have taken some value $y^*\neq y$.

The first condition of *Hitchcock* simply requires cause and effect to be actual events. Or more precisely, the first condition requires that the assignment $\boldsymbol{A}$ and equations $\boldsymbol{E}$ of model $\boldsymbol{M}$ together set X to x and Y to y. The preceding requirement that $\boldsymbol{M}$ be *apt* (more on this in §1.3) then imposes the general requirement that $\boldsymbol{M}$ sets a variable V to value v if and only if $V=v$ correctly characterizes an actual event. After

[5] For various accounts of actual causation within the causal models framework, see Hitchcock 2001, Woodward 2003, Halpern and Pearl 2005, and Weslake forthcoming, *inter alia*.

all, in order to count as apt, the model must match reality with respect to the events it characterizes. In this way, the first condition and the aptness requirement operate together to require that cause and effect be actual events.

The real action in *Hitchcock* is in condition 2, which may be glossed as follows. There should be a path ***P***—a parenthood chain—from X to Y that is *intrinsically right* for counterfactual dependence.[6] What makes a path ***P*** intrinsically right for counterfactual dependence is that there is some way to set the values of the variables off ***P*** under which $Y=y$ counterfactually depends on $X=x$ (the dependence gets revealed when the background conditions are right). More precisely, condition 2 requires that there be a path ***P*** and a possible setting of values to the variables off ***P*** that preserves the values of all variables on ***P*** (except perhaps X), under which setting X to x would set Y to y, while setting X to at least one of its alternative values x^* would set Y to an alternative value y^*. Under the background conditions of this possible setting of values to the off-path variables, 'wiggling' the value of X from x to some other value x^* wiggles the value of Y from y to some other value y^*.

One can put *Hitchcock* to work to get the intuitively plausible result that, in the overdetermination case *Two Rocks*, each throw counts as an individual cause of the shattering. (This is an encouraging result because counterfactual accounts are known to struggle with overdetermination cases.) Assuming that $\boldsymbol{M_2}$ is indeed an apt model, one can use *Hitchcock* to derive that $Throw_1$=1 causes *Shatter*=1 and that $Throw_2$=1 causes *Shatter*=1. Here is how to derive $Throw_1$=1 causes *Shatter*=1 (the case of $Throw_2$ is completely analogous). The first condition is easy: there is an apt causal model $\boldsymbol{M_2}$ such that $Throw_1$=1 and *Shatter*=1. For the second condition, first note that <$Throw_1$, *Shatter*> forms a path $\boldsymbol{P_1}$ with the single off-path variable $Throw_2$, and that there is a possible setting of $Throw_2$ to 0. (One has now pasted in the $Throw_1$ to *Shatter* path into a background setting where no other rock is thrown, in order to find the latent counterfactual dependence there revealed.) With $Throw_2$ set to 0, one finds that *Shatter* still takes its actual value as 2a requires, one finds that 'if $Throw_1$=1 then *Shatter*=1' holds as 2b requires, and one finds that 'if $Throw_1$=0 then *Shatter*=0' holds as 2c requires.

We don't mean to assume that *Hitchcock* is true (in fact *Hitchcock* is widely thought to face counterexamples). Indeed our discussion is complicated by the fact that there is as of yet no consensus on how best to understand actual causation within the causal models framework, and so we need to remain neutral on the matter. We only mean to exhibit one elegant and natural account of actual causation within the causal models framework, as a heuristic for discussing whether causal models need to be supplemented with a default function.

[6] This may be thought of as one precise way to implement Lewis's (1986a: 205–7) vague idea of quasi-dependence.

1.3 Apt Models

One of the key notions that appears in *Hitchcock* and recurs throughout the literature on causal modelling is that of an *apt* (or *fitting*, or *appropriate*) model. And no wonder: causal models are mathematical representations of concrete situations, and whenever one indulges in representation the question may arise as to whether the representation is faithful to reality. This is a very general sort of question arising whenever representations are used (not just in causal modelling). Still, though the question of apt representation arises generally, different uses of representations might call for different sorts of answers, so there may be specific things to say about when representations are specifically apt for causal structure, or even specifically apt for capturing actual causation, or even specifically apt for capturing any actual causal relationships between *this* event and *that* one in light of a particular background inquiry, etc.

There is also a second and largely independent metaphysical question of how to think about actual causation on the (highly plausible) assumption that sometimes *many* causal models are going to pass any plausible test for aptness.[7] What is the best thing to say about actual causation in the world when multiple apt causal models *disagree*? To illustrate the difficulty of this question, suppose that at the end of the day there are exactly two apt models of a given situation, both of which use the variables X, Y, and Z to respectively represent the events *c*, *d*, and *e*. And suppose that one of these two apt models has $X{=}x$ as the actual cause of $Z{=}z$ (and represents no other actual causation), while the other has $Y{=}y$ as the actual cause of $Z{=}z$ (and represents no other actual causation). Now what should one say about the actual causal relationships in the world? *Hitchcock* embeds the existential requirement that there be an apt causal model representing causation, so *Hitchcock* issues the verdict that both *c* and *d* cause *e*. This is surprising: one thing that all of the apt causal models seemed to be agreeing on was that *e* has exactly one cause and not two. One could also embed the universal requirement that every apt causal model represents causation, thereby delivering the verdict that *e* is uncaused. Again this is surprising: one thing that all of the apt causal models seemed to be agreeing on was that *e* has exactly one cause and not none. One might also consider further options including supervaluating over apt models, or saying that whether one event causes another is *relative*

[7] There are at least two respects in which it is hard to imagine the constraints on aptness pinning down a unique causal model. First, there is the question of which events get included. Given the requirements that $\boldsymbol{U}$ and $\boldsymbol{V}$ be finite sets (§1.1), and indeed the incapacity to handle dense causal chains in terms of 'parenthood', it seems that one can only represent a discrete selection of the events involved. It is hard to see an objectively unique determinant of exactly which events must be included and which excluded in all cases. As Halpern and Hitchcock (2010: 394) put the point: 'A modeler has considerable leeway in choosing which variables to include in a model. Nature does not provide a uniquely correct set of variables.' Secondly, there is the question of which contrast possibilities get allotted (cf. Schaffer 2010: §1.3). Given the requirement that $\boldsymbol{R}$ assign each variable $V{\in}\boldsymbol{U}\cup\boldsymbol{V}$ an at-least-two-membered set of options, it seems hard to imagine an objective determinant of exactly which alternatives must be allotted in all cases. Nature does not seem to provide a uniquely correct set of alternatives either.

to a representation, so that the causal facts include the fact that *c* causes *e* relative to the first model, and the fact that *c* does not cause *e* relative to the second model (Halpern and Pearl 2005: 845). For many metaphysicians who think of causation as an objective feature of the natural world, this representation-relativity may be shocking.[8]

We won't have much to say about this second issue, and want to remain neutral on the best thing to say about actual causation when multiple apt models disagree. Our focus is on the prior issue of what makes a model apt in the first place. This is an issue on which there has been comparatively little discussion.[9] Pearl (2000; cf. Hitchcock 2007: 503) relegates the issue to the heading of 'the art' (as opposed to 'the science') of causal modelling, and Halpern and Pearl (2005: 845) speak of this as 'the type of debate that goes on in informal (and legal) arguments all the time', and leave the matter at that. Since the structural equations framework has been developed primarily by statisticians and computer scientists, it is perhaps unsurprising that attention has been focused on developing the mathematical aspects of the models, and drawn away from the metaphysical question of the relation between the model and reality.

Still, one does find some brief guidance in the literature. Hitchcock (2001: 287), for instance, offers the following three necessary conditions on aptness:

1. The counterfactuals encoded in the model's equations must be true;[10]
2. The values of variables should not represent events that bear logical or metaphysical relations to each other;[11]
3. The variables should not be allotted values that one is not willing to take seriously.

Condition 3 plays a role in the discussion to come, and comes in for further clarification and explanation in §3.2. Other natural necessary conditions include:

4. Different values of the same variable should represent noncompossible alternatives;
5. The values allotted should represent intrinsic characterizations;

[8] Perhaps this should not be so shocking. To the extent that there are significant objective constraints on aptness, and to the extent that all of the apt models tend to agree on clear-cut cases, it's not obvious to us that some lingering representation relativity around the margins would be so terrible for actual causation. Indeed, as Halpern and Hitchcock (2010: 384) argue, 'the experimental evidence certainly suggests that people's views of causality are subjective'. See §2.3 and §3.4 for further discussion.

[9] We thus agree with Paul and Hall (2013: 18–19) who say: 'It is an excellent question, inadequately addressed in the literature, precisely what principles should guide the construction of a causal model.'

[10] The extent to which this is an objective constraint corresponds to the extent to which the truth of counterfactuals is an objective matter. But note that constraint 1 embeds a conception of counterfactuals as having model-independent truth values, which we then require apt models to uphold. This means that it is not possible to use causal models to give the semantics for counterfactuals (cf. Shulz 2011; Briggs 2012). Rather we need a prior model-independent semantics for counterfactuals, in order to evaluate model aptness by constraint 1.

[11] This corresponds to the metaphysician's requirement that the events involved be distinct events (Lewis 1986b).

6. The assignment should represent events correctly.[12]

Hitchcock (2007: 503) adds a further condition:

7. The variables should represent enough events to capture the essential structure of the situation being modelled.

Perhaps the most detailed discussion of aptness in the literature is found in Halpern and Hitchcock (2010: §§4–5). Indeed they (2010: 386) specifically worry that, by incorporating default information, 'the problem of justifying the model becomes even more acute'. Halpern and Hitchcock add further necessary conditions on aptness including a condition which is related to condition 7:

8. *Stability*: Adding additional variables should not overturn the causal verdicts.

Conditions 7 and 8 play a pivotal role in the discussion to come, and come in for further clarification and explanation in §3.3.[13]

We would emphasize that all of these are vague conditions, aspects of the art rather than the science of causal modelling. In no case does one find a mathematically precise account of these conditions within the terms of the structural equations framework. Rather these are extra-mathematical conditions on the relation between the mathematics and the reality it would represent. Do not expect more.

Where this is going: We think that some key arguments for default-relativity rely on non-apt models.

2 The Case for Default-Relative Models

> [T]ypical accounts fail to incorporate a distinction between the *default* behavior of an object or system, and *deviations* therefrom...This oversight is fatal; rectify it, and it becomes easy to produce a vastly improved structural equations account. (Hall 2007: 110)

Historically, the idea that judgements of actual causation are relative to some consideration of what counts as a 'default' or normal condition traces back to

[12] In our set-up, apt for the deterministic case, the assignment only covers the initial conditions. But given that the initial conditions are represented correctly (as per 3) and that the counterfactuals encoded in the equations are true (as per 1), the model must in the end represent all events correctly.

[13] We include both conditions 7 and 8 because there are legitimate concerns one might have over each condition individually (it helps buttress our results if both of these conditions converge, even if each condition individually is at best a defeasible heuristic). With 7 there is a concern over vagueness: what is the essential structure of the situation, independent of the models that are supposed to display causal structure? With 8 there are concerns over cases in which one can interpolate variables to overturn good causal verdicts. (There is also a concern about the computational tractability of actually trying to use 8.) In the cases where we apply 7 and 8 we think that there is a core phenomenon of *an impoverished model that omits crucial information*. We think that there needs to be some constraint corresponding to the vague idea of 'don't use impoverished models'. Any such constraint should equally be able to do the work we put 7 and 8 towards.

Mill's (1950) account of our capricious selection of 'the cause' from among the many relevant conditions, and plays a prominent role in Hart and Honoré's (1985) treatment of causal thinking in the law, while continuing to show up in some recent discussions of causation not couched in terms of structural equations, including Maudlin 2004 and McGrath 2005. But the idea that standard causal models need to be supplemented to distinguish default from deviant events—the idea of connecting the normativity of causal judgement to the formalism of causal modelling by supplementing the latter—traces back to Menzies (2004, 2007), and is further developed in Hall 2007, Hitchcock 2007, Halpern 2008, and Livengood 2013. Since Menzies seems to have been a driving force behind this idea we call it in his name:

(*Menzies*) The formalism of standard causal models must be supplemented to distinguish default from deviant events, in order to capture the facts of actual causation.

Menzies is an intriguing and increasingly influential idea. It may well be right (though we are sceptical). We turn to reviewing the case for *Menzies*, which is said to provide a conservative and psychologically plausible extension of standard causal modelling, in ways that solve multiple problems. This is to set the stage for our critique (§3).

2.1 The Gardener and the Queen

One standing problem in treatments of actual causation arises with omissions. Suppose that the gardener is supposed to water the flowers, fails to do so, and the flowers die. People tend to blame the gardener and speak of his failure to water the flowers as a cause of their death. But what is so special about the gardener? The queen of England and indeed everyone else equally failed to water those flowers. Indeed, metaphysically speaking, the gardener and the queen seem perfectly on par: neither actually watered the flowers, and each is such that, counterfactually, if he/she had watered the flowers then the flowers would have survived (Hart and Honoré 1985: 38; Menzies 2004: 145; McGrath 2005; Sartorio 2010).

So is it possible to capture the intuitive conjoined verdict that the gardener's failure to water the flowers caused them to die, but the queen's failure to water the flowers did not? Of course there is room to debate whether this is a verdict worth capturing. Some (such as Beebee 2004) would deny that the gardener's failure to water caused them to die on grounds that absences cannot feature in causal relations at all. Others (such as Lewis 1986a; cf. Schaffer 2004 and 2012) would maintain that the queen's failure to water the flowers did cause them to die, but add some further pragmatic or psychological explanation for why people focus on the gardener rather than the queen. But let us suppose that one wants to capture a distinction in actual causation between the gardener and the queen. How might one do so?

The notion of a default seems tailor-made to capture a gardener/queen distinction. As Hart and Honoré (1985: 38; cf. Menzies 2004: 176) comment, 'The "failure" on the part of persons other than the gardener to water the flowers would...be a normal

though negative condition...The gardener's failure to water the flowers, however, stands on a different footing.' At least in the straightforward versions of the case, the gardener's failure to water the flowers stands out as being in some sense deviant, or abnormal, or unexpected. (Unclarities creep in: what if the gardener is known to be a hopeless shirker who never shows up for the job?) The queen's failure to water the flowers, in contrast, is a default, normal, and expected affair.

So one sees an initial case for *Menzies*, in the form of a problem that default-relativity could potentially solve, namely the problem of drawing a causal distinction (assuming one is wanted) between the gardener and the queen. We should clarify that all we have so far is the possibility of solving a problem (or better: the possibility of drawing a distinction that not everyone agrees should be drawn). One still needs to develop a device to track default versus deviant status in the formalism, and to describe a way to make use of this device in a revised account of actual causation. The prospect of distinguishing the gardener from the queen is one motivation for these developments.

2.2 The Vandal and the Guard (Structural Isomorphs)

Perhaps the strongest argument for default-relativity—or at least the argument that strikes us as being strongest, which Menzies (personal communication) also considers strongest, and which Halpern and Hitchock (forthcoming: §1) identify as a main motivation—comes from Hall's (2007: 121–4) structurally isomorphic but causally distinct cases. Schematically speaking, the argument works by presenting a pair of stories that intuitively differ causally. Causal models of these stories are then presented, and the models turn out to be structurally isomorphic. It is concluded that standard causal models simply cannot see any difference whatsoever between these stories. Default-relativity then is shown to account for the causal difference in a natural way.

As Hitchcock (2007) notes, Hall's original isomorphic cases turn out to be more complicated than the argument requires. The argument can be made by comparing cases of overdetermination (such as the case of *Two Rocks*, §1.1), with the 'bogus prevention' cases introduced by Hiddleston (2005) as counterexamples to *Hitchcock*. So consider the following bogus prevention story:

> Killer plans to poison Victim's coffee, but has a change of heart and refrains from administering the lethal poison. Bodyguard puts an antidote in the coffee that would have neutralized the poison (had there been any present). Victim drinks the coffee and (of course) survives.

A very natural intuition to have about this story is that Bodyguard's putting the antidote in the coffee did not save Victim's life, since there was never any real threat to Victim's life.

Now a seemingly natural minimal way to model this story is by introducing three binary variables:

Poison=1 if Killer does *not* administer the poison, 0 otherwise[14]
Antidote=1 if Bodyguard administers the antidote, 0 otherwise
Survival=1 if Victim survives, 0 otherwise

Poison and *Antidote* are exogenous. *Survival* is endogenous and associated with the equation: *Survival*←max(*Poison*, *Antidote*). Finally one assigns 1 to both *Poison* and *Antidote*. The model is then:

Bogus Prevention
$\boldsymbol{S_{bog}}$ = <{*Poison*, *Antidote*}, {*Survival*}, $\boldsymbol{R_{bog}}$>, where $\boldsymbol{R_{bog}}$ maps all variables to {0, 1}
$\boldsymbol{L_{bog}}$ = <$\boldsymbol{S_{bog}}$, {*Survival*←max(*Poison*, *Antidote*)}>
$\boldsymbol{M_{bog}}$ = <$\boldsymbol{L_{bog}}$, {*Poison*=1, *Antidote*=1}>

But wait! $\boldsymbol{M_{bog}}$ is perfectly isomorphic to the model $\boldsymbol{M_2}$ for overdetermination:

Two Rocks
$\boldsymbol{S_2}$ = <{$Throw_1$, $Throw_2$}, {*Shatter*}, $\boldsymbol{R_2}$>, where $\boldsymbol{R_2}$ maps all variables to {0, 1}
$\boldsymbol{L_2}$ = <$\boldsymbol{S_2}$, {*Shatter*←max($Throw_1$, $Throw_2$)}>
$\boldsymbol{M_2}$ = <$\boldsymbol{L_2}$, {$Throw_1$=1, $Throw_2$=1}>

And moreover it is very natural to think that each throw is a cause of the window shattering in this case. Indeed it was one of the key successes of *Hitchcock* that it got this right (§1.2).

So there seems to be trouble for standard causal models. Since $\boldsymbol{M_{bog}}$ is isomorphic to $\boldsymbol{M_2}$, there seems to be no possibility of distinguishing the causal status of *Antidote* and $Throw_2$ in standard causal models. In this vein Hall (2007: 124; cf. Halpern and Hitchcock forthcoming: §4) claims: '[T]he isomorphism between the models establishes—*conclusively*—that the account...will inevitably be forced to say that the two causal structures are the same. But they aren't. So something has gone badly wrong.' What to do?

Before explaining how default-relativity might help, we first note that this argument is not nearly as conclusive as Hall claims. The appearance of conclusiveness comes from neglecting considerations about how to understand actual causation given the presumptive multiplicity of apt models (§1.3). If $\boldsymbol{M_{bog}}$ and $\boldsymbol{M_2}$ were the one and only apt causal models of their associated situations then the argument would be conclusive. But we know that typically there are many apt models of a given situation, and we are staying neutral between treating actual causation as model-relative (*à la* Halpern and Pearl 2005) or by existentially quantifying over apt models (*à la* Hitchcock 2001), *inter alia*.

[14] Here we depart temporarily from the standard convention of using '1's to represent occurrences and '0's for absences. This is harmless, since these conventions are of course just this: conventions. (We do this just to make the isomorphism with our model for overdetermination more obvious.)

So suppose that one goes in for model-relative actual causation. Then the result is that *Antidote* causes *Survival* relative to $\boldsymbol{M_{bog}}$ if and only if $Throw_2$ causes *Shatter* relative to $\boldsymbol{M_2}$. That result is not obviously so bad. It is perfectly consistent with Bodyguard's putting the antidote in Victim's coffee, and the throwing of the second rock, having very different causal statuses relative to all sorts of further (non-isomorphic) apt models.

Or suppose that one goes in for existential quantification over apt models, but goes in for a more demanding account of actual causation than given in *Hitchcock*, one which does not count $Throw_1$ or $Throw_2$ as actual causes of *Shatter* in $\boldsymbol{M_2}$. Then by isomorphism one likewise gets the result that *Antidote* is not a cause of *Survival* in $\boldsymbol{M_{bog}}$. But that is all fine, and indeed all still comes out well in the end provided that there is some *other* apt model $\boldsymbol{M_3}$ for the two rocks case on which $Throw_1$ does cause *Shatter*, as well as some apt model $\boldsymbol{M_4}$ (which might or might not be identical to $\boldsymbol{M_3}$) on which $Throw_2$ causes *Shatter*, but no apt model on which *Antidote* causes *Survival*. So given existential quantification over apt models, the isomorphism between $\boldsymbol{M_{bog}}$ and $\boldsymbol{M_2}$ is still completely consistent with the desired end result (e.g. that the throwing of the second rock caused the shattering, but that Bodyguard's administering the antidote did not cause the surviving).

But never mind all that. We have a different response to make: we don't think that $\boldsymbol{M_{bog}}$ is an apt model in the first place (§3.3).

Still, just to see the argument through to the point where defaults get introduced, let us follow along (if only to see how much unclarity this talk of defaults brings). Suppose one decides that criminal acts of vandalism like throwing rocks at windows are always deviant. Then both $Throw_1$=1 and $Throw_2$=1 count as deviant in *Two Rocks*. (Is that right? What if vandalism happens a lot? What if the throws aren't acts of vandalism but actually the contractually required efforts of an expert demolition team?) Suppose one also decides that criminal acts like poisonings are deviant but that protective acts like administering antidotes are not deviant but default. (Is that right? What if no one has ever administered antidote before? What if Victim is in line for capital punishment, Assassin is a reluctant executioner who cannot bring himself to follow the law, and Bodyguard is one of Victim's gang attempting to thwart the law?) If so then one has a foothold on distinguishing these cases.

So one sees a second case for *Menzies*, though again we should clarify that all one has so far is the possibility of solving a problem. One still needs to develop a device to track default versus deviant status in the formalism, and to describe a way to make use of this device in a revised account of actual causation. The prospect of distinguishing the vandal from the guard is one more motivation for these developments.

2.3 Psychological Plausibility

Advocates of default-relativity also draw on psychological evidence that our judgements of actual causation are tied into considerations of normality. As Kahneman and Miller (1986: 149)—in a seminal discussion of availability heuristics in the

context of causal and counterfactual judgement—write: 'A cause must be an event that could easily have been otherwise. In particular, a cause cannot be a default value among the elements that the event X has invoked.' Likewise Byrne (2011: 209) notes that our counterfactual reasoning seems biased towards the replacement of abnormal by normal factors: 'People mentally "undo" the exceptional event to make it normal, and they do so regardless of which event in the scenario is indicated to be exceptional.'

Indeed, recent experimental work on causal judgement suggests that moral valence matters to judgements of actual causation. In this vein Hitchcock and Knobe (2009) present a variety of paired cases differing only over the background matter of compliance with norms, and find robust differences in causal judgements, suggesting that judgements of actual causation are serving the practical end of identifying what to fix. As Halpern and Hitchcock (forthcoming: §5; cf. Menzies 2011) put the point: 'Actual causation...is a fairly specialized causal notion. Actual causation involves the *post hoc* attribution of causal responsibility for some outcome. It is particularly relevant to making determinations of moral or legal responsibility.'

And so the friend of *Menzies* may say that our specific concept of actual causation—as distinct from other causal concepts like those given in the structural equations themselves—is a marginal and normatively loaded concept. It is just a tool the folk use for placing blame. Default-relativity starts to look like a psychologically plausible component of such a tool.

2.4 Conservative Extension

So far we have reviewed two things that one can potentially do with a default/deviant distinction: distinguish the gardener from the queen, and distinguish the vandal from the guard. And we have reviewed reasons for thinking that this distinction plays a psychological role in our judgements of actual causation. What makes default-relativity especially promising is that there seems to be a way—indeed at least two ways—to actually do all of this with just a small and conservative extension of standard causal models, which in no way affects the successes of this technology in matters such as causal discovery.

The first strategy we mention—due to Hitchcock (2007: 506–7)—proceeds in two steps. First, it adds an extra element to causal models. This extra element is a function $\boldsymbol{D}$ which takes in the output of the range function $\boldsymbol{R}$ (mapping every variable $V \in \boldsymbol{U} \cup \boldsymbol{V}$ to an at-least-two-membered set of allotted values), and splits this output into default and deviant values. $\boldsymbol{D}$ is required to be a total function on the output of $\boldsymbol{R}$. It must decide on every value of every variable, whether that value is default or deviant. $\boldsymbol{D}$ might or might not also be required, for every variable, to map at least one of its values to default and/or to map at least one of its values to deviant. Finally, $\boldsymbol{D}$ can look at the equations $\boldsymbol{E}$ and the assignment $\boldsymbol{A}$ (so it should only be added after the assignment is in place: §1.1), and can map a given value of endogenous variable $X \in \boldsymbol{V}$ to default or deviant depending on the values that X's parents take. For instance,

if Y is the one and only parent of X, and each has seven allotted values {0...6}, $\boldsymbol{D}$ can treat the default value of X as the actual value of Y (whatever that may be), thus treating X as an 'inertial' variable expected to match the value of Y whatever the value of Y may be.

The entire system of structural equations can then remain unchanged except for some second adjustment to the definition of actual causation to incorporate $\boldsymbol{D}$. (Since one has merely added the extra element $\boldsymbol{D}$, and then only invoked $\boldsymbol{D}$ in the account of actual causation, the extension is completely conservative: every previous application of the structural equations framework remains in place.)

The adjustment to the definition of actual causation could go in different ways, but let us just illustrate one small and crude modification to *Hitchcock* that handles the cases under discussion (there are other problems with this account but they won't be relevant here). This modification is based on a very literal treatment of Kahneman and Miller's (1986: 149) rule that 'a default value cannot be presented as a cause', via a final third condition appended to *Hitchcock* (the first two conditions are unchanged):

(*Hitchcock-meets-Menzies*) $X{=}x$ is an actual cause of $Y{=}y$ if and only if there is an apt causal model $\boldsymbol{M}$ such that:

1. The actual values of X and Y in $\boldsymbol{M}$ are x and y, respectively.
2. There is a directed path $\boldsymbol{P}$ from X to Y and a possible assignment of values $\boldsymbol{z}$ to the set of variables $\boldsymbol{Z}$ off $\boldsymbol{P}$ such that the following counterfactuals are true:
   a. Had $\boldsymbol{Z}$ taken values $\boldsymbol{z}$, the variables on $\boldsymbol{P}$ (except possibly X) would still have taken their actual values.
   b. Had $\boldsymbol{Z}$ taken values $\boldsymbol{z}$ and X value x, Y would have taken value y.
   c. Had $\boldsymbol{Z}$ taken values $\boldsymbol{z}$ and X some value $x^{*}{\neq}x$, Y would have taken some value $y^{*}{\neq}y$.
3. $X{=}x$ is a deviant value

In one fell swoop, *Hitchcock-meets-Menzies* promises to distinguish the gardener from the queen and the vandal from the guard. Or at least, given that the default function $\boldsymbol{D}$ assigns the gardener's failing to water the flowers to *deviant* but the queen's failing to water the flowers to *default*, it only allows the former to be causal. And given that $\boldsymbol{D}$ assigns the vandal's throwing the rock as *deviant* but the bodyguard's administering the antidote as *default*, it likewise only allows the former to count as causal. And it does so in a way that honours Kahneman and Miller's dictum, without upsetting anything else in the structural equations framework. Not bad.

The second strategy we mention—pursued by Menzies (2007)—also proceeds in two steps (both of which are different from the steps taken by the first strategy). Going back to standard causal models, the second strategy adds a *normality ranking function* $\boldsymbol{N}$ over sets of possible total states of the system, where a possibly total state of the system is a complete valuation of all variables consistent with the structural

equations (in the deterministic case we are assuming here, these are given by all the possible assignments a given linkage can bear).[15] For instance, *Two Rocks* has four possible total states:

State1: $Throw_1$=1, $Throw_2$=1, *Shatter*=1
State2: $Throw_1$=1, $Throw_2$=0, *Shatter*=1
State3: $Throw_1$=0, $Throw_2$=1, *Shatter*=1
State4: $Throw_1$=0, $Throw_2$=0, *Shatter*=0

(These states correspond to the four possible assignments given two binary exogenous variables, together with the value of the one endogenous variable as fixed by the equations.) Given that vandalism counts as abnormal (and each throw equally so), the ranking from most normal (rank 0) to least might run:

N_2: State4 → 0 (no vandalism so most natural)
State3, State2 → 1 (one act of vandalism so tied for second)
State1 → 2 (two acts of vandalism so least natural)

So instead of splitting individual values of variables into default/deviant as with *Hitchcock-meets-Menzies*, now one is ranking total states of the system for normalcy.[16]

One could now modify the definition of actual causation to incorporate ***N***, but we instead follow Menzies (2004, 2007, 2009, 2011) and incorporate ***N*** in a deeper way, into the evaluation of counterfactuals themselves within the structural equations framework. (This might seem riskier, insofar as now one risks upsetting other applications that use counterfactuals. But keep in mind that one can always keep two notions of a 'counterfactual' in this framework, and reserve the new method of evaluation just for actual causation or some other limited range of applications. One does not lose access to the old technique for evaluating counterfactuals, just by introducing a new one.)

The evaluation of counterfactuals can be modified in various ways, but let us just illustrate one small modification that handles the cases under discussion, based on a very literal treatment of Hart and Honoré's (1985: 29; cf. Menzies 2009: 355–6) idea of the cause as 'a difference from the normal course':

(*Menzies-by-Menzies*) $X=x$ is an actual cause of $Y=y$ if and only if there is an apt causal model ***M*** such that:

[15] One might also employ the weaker notion of a partial ordering of possible total states of the system, and/or rank all total states of the system including those incompatible with the equations (Halpern and Hitchcock forthcoming).

[16] Of course the normality of individual values of variables and the normality of total states of the system are not unrelated matters. But there needn't be any simple correspondence either. Total normality might be a complicated and holistic affair.

1. The actual values of X and Y in $\boldsymbol{M}$ are x and y, respectively.
2. Had X taken some value $x^{*}\neq x$, Y would have taken some value $y^{*}\neq y$.[17]

Menzies-by-Menzies reverts to a very direct association between causation and counterfactual dependence. Gone are the clauses of *Hitchcock* that identify directed paths and consider alternative settings for off-path variables.

The main novelty in *Menzies-by-Menzies* comes in the interpretation of the counterfactual in 2, which is evaluated by a recipe that begins, in a new way, by moving out of the actual assigned model to the set of most normal counterparts:

> *Re-centre on the most normal states*: Consider the set $\boldsymbol{M_{def}}$ of assigned models that have the same linkage as $\boldsymbol{M}$, but an assignment $\boldsymbol{A_{def}}$ that induces one of the most normal states (according to the ranking in $\boldsymbol{N}$).

We then run our old-style evaluation of counterfactuals from each model in $\boldsymbol{M_{def}}$: we cut any incoming links to X and 'surgically' reassign alternative values, and look to see whether the value of Y is any different (§1.2). What is happening is that we are not evaluating the counterfactual directly at the actual world, but evaluating it by centring on the most normal worlds instead, and treating causation as counterfactual dependence at the most normal worlds.

There are problems with this simple approach but they won't be relevant here. Let's just look at how this approach enjoys some successes. Starting with the gardener and the queen, we start not in the actual state but in the most normal state, which is taken to be one in which the gardener waters the flowers, the queen does not, and the flowers survive. Relative to that normal state, a modification of what the gardener does 'wiggles' the state of the flowers (if the gardener fails to water the flowers, they die), but a modification of what the queen does makes no difference to the flowers (they survive regardless, thanks to the gardener). So we get the gardener but not the queen as a cause. With the vandal, we start in the most normal state, which is taken to be State4 above, in which neither rock is thrown and the window remains intact. Relative to State4, a modification of either throw 'wiggles' the state of the window (if either vandal throws, the window shatters). So we get both vandals as causes. And finally, with the guard, we start in the most normal state, which is taken to be one where Assassin does not put poison in the coffee and Victim survives. We don't even need to decide whether Bodyguard administers the antidote or not in the most normal state (we can call it a tie if we like). For as long as Assassin does not put poison in the coffee, a modification of what the guard does makes no difference to

[17] Menzies himself—like Hitchcock (1996) and Schaffer (2005), *inter alia*—is a contrastivist about causation, so strictly speaking (for fidelity) we should be talking about $X{=}x$ rather than x^{*} causing $Y{=}y$ rather than y^{*}, and invoking this specific contrast in clause 2. For present purposes we suppress this complication. We are not trying to get the full details of Menzies' positive approach on the table, but just display a way to put defaults to work.

Victim (there is no threat so Victim survives regardless). So, while we got both vandals as causes, we do not get the guard as a cause.

Thus both *Hitchcock-meets-Menzies* and *Menzies-by-Menzies* do fairly well at least in the hard cases considered, while honouring some plausible background ideas, and without upsetting anything else in the structural equations framework. Not bad at all. Maybe these ideas are moving in the right direction?[18]

3 The Case Against Default-Relativity

> [W]hat makes a model an appropriate model? While we do want to allow for subjectivity, we need to be able to justify the modelling choices made. A lawyer in court trying to argue that faulty brakes were the cause of the accident needs to be able to justify his model; similarly, his opponent will need to understand what counts as a legitimate attack on the model.
>
> (Halpern and Hitchcock 2010: 384–5)

While we are open to default-relative models as per *Menzies*, we have concerns, and are largely unmoved by the case just sketched (§2). We think that both the argument about being able to distinguish the gardener and the queen, and the argument about being able to distinguish the vandal and the guard, rely on non-apt models. What is to be learned from these arguments is not that causal models must be default-relative, but that more attention is needed concerning what counts as an apt causal model in the first place.

3.1 Unclarities

What concerns us most about default-relativity is the unclarities it generates (some of which have already been noted). Normality judgements seem to draw on diverse and typically competing factors, in a highly context-sensitive way. As such default-relativity often seems to us to come close to a free parameter in an otherwise so precise and objectively constrained formalism, which basically gives the theorist leeway to hand-write the result she wants. So we think that default-relativity generates complicating and under-constrained unclarities.

Advocates of default-relativity have made some efforts to sketch constraints. For instance, Menzies (2011: 196–7), drawing on the psychological literature on counterfactual availability, says:

> [A] normal event or state of affairs is one that is common, expected, and unsurprising, whereas an abnormal event or state of affairs is one that is exceptional, unexpected, and

[18] Both *Hitchcock-meets-Menzies* and *Menzies-by-Menzies* add default information into the internal mathematical structure of causal models. It is worth noting a distinct (though compatible) way in which defaults might be used, namely as providing external constraints on model aptness. For instance, one could claim that, for a model to be apt, the assignment should set no exogenous variables *to default values*, perhaps on grounds that any exogenous variable set to a default value represents a 'background condition' that ought to be relegated to what Mackie (1974) calls 'the causal field' rather than included within the model.

surprising...[A] normal event is one that conforms to the norms and an abnormal one is one that violates the norms, where the relevant norms can be evaluative or empirical.

We find expressions like 'a normal event is one that conforms to the norms' to be unhelpful, but we note the helpful specification of evaluative and empirical components of normality, which suggests moral permissibility and statistical likelihood as components of normality. But even then we get no guidance for cases in which these components conflict. What counts as 'normal' in a den of thieves, where the morally permissible is statistically unlikely?

Halpern and Hitchcock (2010: 402–3) go on to specify four components of normality:

- Statistical norms concerning what happens most frequently
- Moral norms concerning what is permissible
- Social norms concerning what policies are in force
- Functional norms concerning how systems are supposed to operate ('there are certain ways that hearts and spark plugs are "supposed" to operate')

This is helpful as well. Yet they explicitly allow that normality may have further components, and give no guidance for typical cases in which these components conflict. What counts as normal if most spark plugs are defective? What counts as normal if a social policy mandates immoral behaviour, or unlikely behaviour? To take a concrete case relevant to the law and public policy, most people speed. If the posted speed limit is 55 miles per hour, is driving at 55mph normal for conforming to the law, or abnormal for violating the statistical expectation? (And if the issue comes before the courts, do we really expect the courts to decide on the matter of 'normality', or to care?)

So we note two respects in which we find default-relativity unclear. The first respect is that default status seems *underdetermined* in many cases. Suppose that the referee flips a coin to determine whether the red team or the blue team gets to kick off a friendly soccer match. The coin lands heads and so red kicks off. This can be naturally modelled with two variables:

Flip = 1 if the coin lands heads, 0 if it lands tails;
Kickoff = 1 if the red team kicks off, 0 if the blue team kicks off.

We get the following model, isomorphic to *One Rock* (§1.1):

Coin Flip
$\boldsymbol{S_{coin}} = \langle\{Flip\}, \{Kickoff\}, \boldsymbol{R_{coin}}\rangle$, where $\boldsymbol{R_{coin}}$ maps both *Flip* and *Kickoff* to {0, 1}
$\boldsymbol{L_{coin}} = \langle \boldsymbol{S_{coin}}, \{Kickoff \leftarrow Flip\}\rangle$
$\boldsymbol{M_{coin}} = \langle \boldsymbol{L_{coin}}, \{Flip=1\}\rangle$

Graphically:

But how are we supposed to assign default or deviant values to events? Statistically speaking we can imagine (and indeed presumably expect) that the frequencies of heads and tails landing are roughly equal (in any reasonable reference class). There is no moral or social preference for heads over tails, and the coin is not 'supposed' to land heads any more than tails. If we have to sort the heads and tails outcomes into default versus deviant, it is hard to see how we could even begin.[19] Likewise it seems hard to assign either value of *Kickoff* (red team versus blue team kicks off) to default or deviant status. (Would it matter if the red team had kicked off in two of the last three meetings, or if the modeller was rooting for red?)

Tellingly, judgements of actual causation seem to us to remain clear and robust even in such cases. In *Coin Flip*, the outcome of the flip (heads versus tails) determines who kicks off (the red team versus the blue team). If one flip outcome and/or one side kicking off is especially 'deviant', the matter seems causally inert. We may ignore it in causal judgement.

The friend of default-relativity might reply that there is a clear verdict on defaults: *a tie*. Friends of a default-function (as per the *Hitchcock-meets-Menzies* proposal in §2.4) could say that there are four apt models into which *Coin Flip* may be extended, differing over the assignment of default/deviant status to *Flip* and *Kickoff*. Relatedly friends of a normality ranking (as per the *Menzies-by-Menzies* proposal in §2.4) could say that there is a single apt model featuring a tie between the two most natural total states (S1: *Flip*=0 and *Kickoff*=0, and S2: *Flip*=1 and *Kickoff*=1). We cannot rule out this reply. Rather our complaint is that we have little idea how to assess it: this is a respect in which the introduction of default-relativity has come at the cost of clarity.

The second respect of unclarity, already mentioned above, is that default status seems *conflictingly overdetermined* in many cases. Suppose—to extend the speeding case above—that Sally (like most other drivers on the road with her) is driving 65mph in a 55mph zone, and gets into an accident that would not have occurred had Sally been driving at 55mph. This can be naturally modelled with two variables:

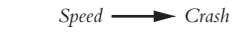
Speed = 1 if Sally drives 65mph, 0 if she drives 55mph
Crash = 1 if Sally gets into an accident, 0 if no accident

We get the following model, isomorphic to both *One Rock* and *Coin Flip*:

Car Crash
$\boldsymbol{S}_{crash}$ = <{*Speed*}, {*Crash*}, $\boldsymbol{R}_{crash}$>, where $V \in \boldsymbol{U} \cup \boldsymbol{V}$ $\boldsymbol{R}_{crash}$ maps both *Speed* and *Crash* to {0, 1}
$\boldsymbol{L}_{crash}$ = <$\boldsymbol{S}_{crash}$, {*Speed*←*Crash*}>
$\boldsymbol{M}_{crash}$ = <$\boldsymbol{L}_{crash}$, {*Speed*=1}>

[19] Though see Diaconis, Holmes, and Montgomery 2007 for an argument that flipped coins show a statistical bias (.51) for landing as they started. Does this finding change the verdict?

Graphically:



But how are we supposed to assign default or deviant values to *Speed*? The social norm is to drive 55mph, but the statistical norm is to drive 65mph. What is morally permissible in this case depends partly on the driving conditions: driving 65mph may even be the only way for Sally to uphold her duty of care to others, given that she is surrounded by fellow speeders. And the functional norms may depend on what model of car Sally is driving: sports cars are supposed to go fast. Should the verdict on actual causation in this case really be sensitive to what model of car Sally is driving? What is normal?

Tellingly, judgements of actual causation seem to us to remain clear and robust even in such cases. In *Car Crash*, Sally's speeding 'wiggles' whether or not she crashes. If one way of driving is on balance 'deviant', the matter seems causally inert. We may ignore it in causal judgement.

Friends of default-relativity could call it a tie (either by using a default-function and going with multiple apt models, or by using a normality ranking with a tie at the top). Though it seems overly crude to call every case in which norms conflict a tie. That approach makes for too many ties in flawed worlds like ours. And it seems to miss the point that a lot of these considerations just don't matter to our causal judgements.

Our underlying point: When we want to model cases like *Coin Flip* and *Car Crash*, things go *much more smoothly* if we don't have to bother with default-relativity. We just follow the simple and elegant treatment of the standard causal modeller, and recover the obvious verdicts of actual causation (on every plausible way of reading actual causation off standard causal models). 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.

3.2 The Gardener and the Queen Revisited

Did we need defaults to distinguish the gardener from the queen with respect to actual causation? For present purposes let us grant the controversial claim that a metaphysical distinction is wanted (§2.2). Instead we propose to argue, first, that defaults don't help, and second, that what does help is to dismiss models that represent the possibility of the queen watering the flowers as non-apt. Such models represent a possibility that we are not willing to take seriously.

Our first argument, to the conclusion that defaults don't help, comes from the claim that the gardener/queen causal asymmetry survives modifications that make the gardener's failure to water the flowers count as 'normal', at least by every single

one of Halpern and Hitchcock's four components of normality (§3.1). Consider a modified version of the case including a little-known secret society called *Flora For Food (FFF)*. *FFF* despises inedible flora (their pamphlets include sections like 'Who eats roses?' and 'Tulips bloom while children starve'). *FFF* members swear the most sacred oaths never to water flowers, and they always keep this oath. *FFF* moreover has infiltrated all levels of government and enacted secret legislation invalidating any contractual requirements for watering flowers. Finally, the gardener is not merely a member of *FFF* but in fact the founder of the whole movement, whose mission in life is to further the cause of *FFF*.

Now we submit that—even if it turned out that the gardener was secretly the founding member of *Flora For Food*—he would still be causally faulted for the death of the flowers, not the queen. Indeed were it discovered that the gardener was secretly a member of *FFF*, the natural reaction would be along the lines of, 'Ah, now we know why he let the flowers die!' rather than 'Oh, now we see that he had no connection at all with the death of the flowers, and was as removed from the case as the queen of England.'

But if the gardener was secretly a member of *Flora For Food*, then his failure to water the flowers counts as 'normal', at least by every single one of Halpern and Hitchcock's four components of normality. The statistical likelihood that the gardener will break his oath and water the flowers is low.[20] Morally, watering the flowers would violate a sacred oath and so would be impermissible. Socially, the success of *FFF* in infiltrating the government has removed any contractual obligation for the gardener to water the flowers. And functionally, given that the gardener's mission in life is furthering the cause of *FFF*, his function does not lie in helping flowers.

(There are moves the friend of *Menzies* could make. Perhaps the standards relative to which the normalcy of the gardener's behaviour is to be evaluated are those that hold in our world, where gardeners routinely water plants, are morally permitted and contractually obligated to do so, and thereby fulfil their functions. Far from alleviating our concerns, the availability of such moves just provides further evidence for our underlying point: the rules for assigning defaults are massively underdetermined.)

So we submit that in our revised version of the gardener case—call it *Secret Gardener*—our intuitions of actual causation still uphold the gardener/queen causal asymmetry, but the status of the gardener's failure to water the flowers seems to count as *default* (just like the queen). So we conclude, on the basis of *Secret Gardener*, that if there is a metaphysical distinction to be drawn between the gardener and the queen, it is not one based on the default/deviant distinction. (We are about to suggest a different basis for the distinction, based on which possibilities we take seriously.)

[20] Of course if one is evaluating frequencies then one needs to specify reference classes. In the main text we are operating with the reference class of the behaviours of this particular gardener. One could instead look to the reference class of gardeners generally, and then matters would turn on how many gardeners were signed on to *FFF*. In the end this is yet another complicating unclarity.

Our second argument is that the gardener/queen causal asymmetry is in any case better treated through constraints on what counts as an apt model. Indeed we recall a standard constraint mentioned in §1.3:

3. The variables should not be allotted values that we are not willing to take seriously

We pause to flag three preceding uses of 3 in the literature (we are not trying to explain the full details, but merely to point the reader to applications). First, Hitchcock—who introduces the constraint in his 2001—uses it to block problematic models of Hall's (2000) boulder cases, where the problematic models allow us to consider the prospect of an overhanging boulder never falling and yet plunging within a metre of Hiker's head (by teleporting?). Constraint 3 also plays a role in Halpern and Pearl's (2005: 877) treatment of McDermott's (1995) fielder/wall asymmetry, based on the idea that we might take seriously the prospect of the fielder failing to catch the ball but not take seriously the prospect of the wall failing to stop the ball. Finally 3 plays *exactly the role we are about to use it for* in Halpern and Pearl's (2005: 871; cf. Woodward 2003: 88) treatment of the gardener/queen asymmetry: 'A model which does not allow us to consider [the queen of England] watering the plant can be defended in the obvious way: that it is a scenario too ridiculous to consider.'

When people intuit that the queen's failure to water the flowers does not cause them to die, they often (if asked) explain themselves by saying something like: 'Come on, the queen of England? She has nothing to do with any of this. She's not going to water the flowers!' To the extent that this is the sort of consideration driving intuitions, it seems that the gardener/queen causal asymmetry is an asymmetry in which possibilities we are willing to take seriously. It is because we are willing to indulge in the fantasy of the gardener watering the flowers (even in *Secret Gardener*), but just can't seriously imagine the queen stooping to the job, that we feel an asymmetry. If so then constraint 3 on apt models—which does independent work—was all we needed to explain the gardener/queen asymmetry. There is no apt causal model in which wiggling whether the queen waters the flowers wiggles the fate of the flowers, because there is no apt causal model that considers so ridiculous a scenario as the queen of England popping by, watering can in hand, to engage in random acts of gardening.

To put this point in another way: the 'problem' crucially relies on using a causal model that represents the queen watering the flowers. But this model is non-apt, for violating an independently justified constraint.

(For those of who think that the gardener/queen asymmetry is not a metaphysical asymmetry but a merely psychological asymmetry, this same underlying point might be put a bit differently. On this way of thinking, the gardener and the queen are equally causes and the models all count as apt. But there is a psychological story to be told about why people tend to focus on the gardener rather than the queen, because

people tend not to consider the model in which the queen is doing anything so inconceivable as stooping to water the flowers. Constraint 3 may be reinterpreted, not as an aptness constraint on models, but as a descriptive psychological claim about which causal models are most readily available to us when we form our causal judgements.)

3.3 The Vandal and the Guard Revisited

Did we really need defaults to distinguish the vandal from the guard with respect to actual causation? Partly in the same way we discussed the gardener/queen asymmetry (which can also be considered a case of isomorphic models but divergent causal judgements), we argue that defaults don't help with the full isomorphism problem, and that what does help with this problem is to dismiss models that fail to represent enough events to bring out the essential causal structure of the situation (and which are failing to issue stable causal verdicts). The problem only arises from the use of impoverished models.

3.3.1 ISOMORPHISM BY IMPOVERISHMENT

Our first argument that defaults don't help comes from considering other paired cases that are structurally isomorphic but causally distinct. If one hunts for isomorphisms (without regard to aptness) they are easy enough to find. Thus consider a paradigmatic *noncause*. For instance, suppose that the innocent bystander watches from her window, gaping in horror, as a child dashes out onto the street and is tragically hit by a car. (There was nothing the bystander could have done to help.) A very minimal but seemingly apt model of this absence of causation is given by the following 'inert' model with just two exogenous variables:

Watch = 1 if the bystander watches, 0 if the bystander looks away
Hit = 1 if the child is hit by the car, 0 if the child is not hit

Innocent Bystander

$\boldsymbol{S}_{bystander}$ = <{*Watch*, *Hit*}, {}, $\boldsymbol{R}_{bystander}$>, where $\boldsymbol{R}_{bystander}$ maps both *Watch* and *Hit* to {0, 1}
$\boldsymbol{L}_{bystander}$ = <$\boldsymbol{S}_{bystander}$, {}>
$\boldsymbol{M}_{bystander}$ = <$\boldsymbol{L}_{bystander}$, {*Watch*=1, *Hit*=1}>

Graphically we get a disconnected structure:

Watch *Hit*

This seems like a reasonable representation of the absence of causal connection between whether or not the innocent bystander watches or looks away, and whether or not the child is hit by the car.

Now for the structurally isomorphic but causally distinct counterpart to pair with *Innocent Bystander*: consider a standard case of early pre-emption. Imagine that the

first vandal is about to throw a rock through the window when she sees the second vandal in action, and so instead just watches as the second vandal throws a rock through the window. Everyone should agree that the second vandal's throw causes the window to shatter. This is a clear instance of causation. But now consider a very minimal model, where one only includes variables for whether or not the second vandal throws, and whether or not the window shatters. Note that (due to the presence of the first vandal, who is still really there but merely left off the model) there is no counterfactual dependence present in reality between the second vandal's throwing and the window's shattering. So aptness condition 1 ('The counterfactuals encoded in the model's equations must be true': §1.1) precludes one from including *any* structural equation linking these variables. The result is that the only possible model apt by the lights of condition 1 is going to be another inert model:

Throw = 1 if the second vandal throws, 0 if the second vandal does not throw
Shatter = 1 if the window shatters, 0 if the window remains intact

Early Pre-emption (Impoverished)
$\boldsymbol{S}_{\textit{early-}}$ = <{*Throw*, *Shatter*}, {}, $\boldsymbol{R}_{\textit{early-}}$>, where $\boldsymbol{R}_{\textit{early-}}$ maps both *Throw* and *Shatter* to {0, 1}
$\boldsymbol{L}_{\textit{early-}}$ = <$\boldsymbol{S}_{\textit{early-}}$, {}>
$\boldsymbol{M}_{\textit{early-}}$ = <$\boldsymbol{L}_{\textit{early-}}$, {*Throw*=1, *Shatter*=1}>

Graphically we again get a disconnected structure:

Throw *Shatter*

And of course $\boldsymbol{M}_{\textit{bystander}}$ and $\boldsymbol{M}_{\textit{early-}}$ are structurally isomorphic.

So we see a further instance of structurally isomorphic but causally distinct cases. But default-relativity seems to have nothing whatsoever to do with the problem of distinguishing bystanders from early pre-emptors. The problem is rather that *Early Pre-emption (Impoverished)* is (as its label indicates) an impoverished model of early pre-emption. One has only found an isomorphism by ignoring essential causal structure, namely the presence of the first vandal. Likewise—under any decent account of actual causation adequate to handle early pre-emption—the causal verdict issued by *Early Pre-emption (Impoverished)* is unstable. Augment the model to include the first vandal (adjusting the remainder accordingly) and the causal connection then appears. To use the language of Halpern and Hitchcock (2010: 384–5), a lawyer seeking to defend the second vandal via *Early Pre-emption (Impoverished)* would be subject to a devastating objection, of the form 'you've missed an essential part of the story, namely that your client acted in the presence of another would-be vandal; consider this and we find the opposite result'.

For a second pair of cases that exhibit the same pattern of isomorphism by impoverishment, start by considering an overdetermining cause as in *Two Rocks*. Pair this with a standard case of late pre-emption in which both vandals throw their rocks but the first vandal's rock arrives first and shatters the window, while the

second vandal's rock arrives second and merely flies through the space where a window recently stood. People tend to think that the overdetermining vandal and the late-pre-empted backup vandal are causally distinct: only the former's actions stand in the actual causation relation to the window shattering. But one can try to model the late pre-emption case with just three binary variables: one for whether or not the first vandal throws, one for whether or not the second vandal throws, and one for whether or not the window shatters. Again aptness condition 1, requiring true counterfactuals, is going to uniquely determine the structural equations. The result is a model isomorphic to overdetermination:

Late Pre-emption (Impoverished)

$\boldsymbol{S}_{late\text{-}} = \langle\{Throw_1, Throw_2\}, \{Shatter\}, \boldsymbol{R}_{late\text{-}}\rangle$, where $\boldsymbol{R}_{late\text{-}}$ maps all variables to {0, 1}
$\boldsymbol{L}_{late\text{-}} = \langle\boldsymbol{S}_{late\text{-}}, \{Shatter \leftarrow \max(Throw_1, Throw_2)\}\rangle$
$\boldsymbol{M}_{late\text{-}} = \langle\boldsymbol{L}_{late\text{-}}, \{Throw_1=1, Throw_2=1\}\rangle$

Again the problem has nothing to do with default-relativity. The problem is rather that *Late Pre-emption (Impoverished)* is (as its label indicates) an impoverished model of late pre-emption. One has only found an isomorphism with overdetermination by ignoring essential causal structure, namely whether each rock actually struck the window. Likewise—under any decent account of actual causation adequate to handle late pre-emption—one finds that the causal verdict issued by *Late Pre-emption (Impoverished)* is unstable. Augment the model to include the extra structure (adjusting the remainder accordingly) and the causal connection from the late-pre-empted backup then disappears. To use the language of Halpern and Hitchcock again, a lawyer seeking to prosecute the late-pre-empted backup vandal via *Late Pre-emption (Impoverished)* would be subject to a devastating objection, of the form 'you've missed an essential part of the story, namely that my client's rock never touched the window; consider this and we find the opposite result'.

3.3.2 BOGUS PREVENTION AS IMPOVERISHED

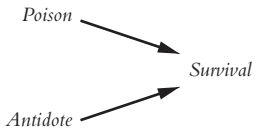
We are now ready to return to *Bogus Prevention*. We think that the treatment of *Bogus Prevention* given in §2.3 is likewise impoverished. That model ignores essential causal structure, namely *whether or not the antidote ever neutralized any poison*. Likewise one finds—using *Hitchcock* to test for actual causation—that the verdict issued by *Bogus Prevention* is unstable. Augment the model to include the extra structure (adjusting the remainder accordingly) and the causal connection from the guard to the survival then disappears.

Let us put this on display. Here is the old model we are claiming is impoverished:

Bogus Prevention

$\boldsymbol{S}_{bog} = \langle\{Poison, Antidote\}, \{Survival\}, \boldsymbol{R}_{bog}\rangle$, where $\boldsymbol{R}_{bog}$ maps all variables to{0, 1}
$\boldsymbol{L}_{bog} = \langle\boldsymbol{S}_{bog}, \{Survival \leftarrow \max(Poison, Antidote)\}\rangle$
$\boldsymbol{M}_{bog} = \langle\boldsymbol{L}_{bog}, \{Poison=1, Antidote=1\}\rangle$

Graphically this looks of course just like overdetermination:



But now we include a binary endogenous variable *Neutral* for whether or not any neutralization occurs, associated naturally with the equation that maps *Neutral* to 1 if and only if *Poison* is at 0 (remember that *Poison*=1 is representing the assassin's *not* administering the poison) and *Antidote* is at 1. That is, the neutralization occurs if and only if both the poison and the antidote are present. So we reach:

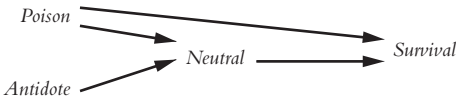
Bogus Prevention (Enriched)

$\boldsymbol{S_{bog+}}$ = <{*Poison*, *Antidote*}, {*Neutral*, *Survival*}, $\boldsymbol{R_{bog+}}$>, where $\boldsymbol{R_{bog+}}$ maps all variables to {0, 1}

$\boldsymbol{L_{bog+}}$ = <$\boldsymbol{S_{bog+}}$, {*Neutral*←min(1-*Poison*, *Antidote*), *Survival*←min(1-*Poison*, Neutral)}>

$\boldsymbol{M_{bog+}}$ = <$\boldsymbol{L_{bog+}}$, {*Poison*=1, *Antidote*=1}>

Graphically the structure is quite different, and looks nothing like simple overdetermination (in fact the model looks more like a standard non-impoverished model for early pre-emption):



More crucially, *Bogus Prevention (Enriched)* reverses the verdict that *Antidote*=1 causes *Survival*=1 (at least continuing to judge these matters by *Hitchcock*). Condition 2 of *Hitchcock* goes unsatisfied. We find a single directed path $\boldsymbol{P_A}$=<*Antidote, Neutral*, *Survival*> running from *Antidote* to *Survival*, with *Poison* being the only variable off $\boldsymbol{P_A}$. There are two possible assignments of values to *Poison* to consider, namely 0 and 1. At possible assignment *Poison*=0 condition 2a fails: this is the case where both the poison and the antidote are administered, and so in this case the variable *Neutral* which lies along $\boldsymbol{P_A}$ would have taken the value 1 and not the actual value 0. At possible assignment *Poison*=1 condition 2c fails: this is the case where there is no poison administered and so we see no way to wiggle the value of *Antidote* in any way that makes a difference to the value of *Survival*. Thus either way condition 2 fails, which delivers the sought verdict that *Antidote*=1 does *not* cause *Survival*=1 in *Bogus Prevention (Enriched)*.

Thus *Bogus Prevention* is ignoring essential structure and issuing an unstable verdict. Just like *Early Pre-emption (Impoverished)* and *Late Pre-emption (Impoverished)* it is violating aptness constraints 7 ('The variables should represent enough events to

bring out the essential causal structure of the situation') and 8 ('*Stability*: Adding additional variables should not overturn the causal verdicts'). To use the language of Halpern and Hitchcock again, imagine that Bodyguard is contractually due a major reward if he saves Victim's life, and tries to claim the reward in court. A lawyer seeking to represent the guard via *Bogus Prevention* would be subject to a devastating objection, of the form 'you've missed an essential part of the story, namely that your client's antidote never actually neutralized any poison; consider this and we find the opposite result'.[21]

3.3.3 EXTENSIONS (BOGUS ANTIDOTE)

Other cases of structurally isomorphic but causally distinct pairs may call for different approaches (there are many ways to make a non-apt model), and may also require revisions to the account of actual causation (we are only using *Hitchcock* as a rule of thumb apt for the cases we are discussing). But we think that—at least in the other isomorphism cases in the literature—winnowing out impoverished models plays a role in the solution. By way of illustration, we extend our treatment to one further pair, which includes the core component of Hall's original pair. Thus consider the following variant on the bogus prevention story (cf. Hitchcock 2007: 519):

> Bodyguard accidentally spills some antidote into Victim's coffee. Killer watches this from a hiding place behind the curtains. Killer is a hired assassin who is being paid to put poison in Victim's coffee. But Killer has also had a change of heart and decided not to kill Victim. Killer was about to simply sneak away, but on seeing Bodyguard spill the antidote into Victim's coffee, Killer sees a perfect solution: he can now put the poison into Victim's coffee (as per his job) without killing Victim (as per his newfound conscience). So Killer waits for Bodyguard to leave the room, adds the poison to the coffee (which is promptly neutralized by the antidote). Victim drinks the coffee and (of course) survives.

A very natural intuition to have about this story is that Bodyguard's putting the antidote in the coffee did not save Victim's life, since the only threat to Victim's life came about through that very act. Had Bodyguard not put the antidote in the coffee there would not have been any threat to Victim at all. (In Hall's terms, Bodyguard's action *short-circuits*: it creates a threat along one path but cancels it out along another.)

Now a seemingly natural minimal way to model this story is by introducing three binary variables akin to the ones used in *Bogus Prevention*:

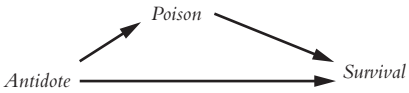
[21] Halpern and Hitchcock (forthcoming: §7.4) note that the extended model *Bogus Prevention (Enriched)* can deliver the right verdict on the case, 'without appeal to normality'. What we are adding: this fact may be exploited to show that the original model of *Bogus Prevention* is non-apt.

Antidote=1 if Bodyguard administers the antidote, 0 otherwise
Poison=1 if Killer does *not* administer the poison, 0 otherwise
Survival=1 if Victim survives, 0 otherwise

Antidote is exogenous. *Poison* is endogenous and associated with the equation: *Poison*←1-*Antidote*. *Survival* is endogenous and associated with the equation: *Survival*←max(*Poison*, *Antidote*). Finally one assigns 1 to *Antidote*. The model is then:

Bogus Antidote
$\boldsymbol{S_{anti}}$ = <{*Antidote*}, {*Poison*, *Survival*}, $\boldsymbol{R_{anti}}$>, where $\boldsymbol{R_{anti}}$ maps all variables to {0, 1}
$\boldsymbol{L_{anti}}$ = <$\boldsymbol{S_{anti}}$, {*Poison*←1-*Antidote*, *Survival*←max(*Poison*, *Antidote*)}>
$\boldsymbol{M_{anti}}$ = <$\boldsymbol{L_{anti}}$, {*Antidote*=1}>

Graphically we get:



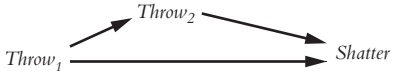
There are now two connected problems. One problem—not crucial given our current concerns—is that *Hitchcock* counterintuitively deems *Antidote*=1 a cause of *Survival*=1. (This is because Hitchcock allows us to freeze *Poison*=0 in considering the efficacy of *Antidote*=1 along the direct path <*Antidote*, *Survival*>, and because at the setting *Poison*=0 we do see counterfactual dependence of *Survival*=1 on *Antidote*=1.) Assuming that one should not count *Antidote*=1 as a cause of *Survival*=1, *Hitchcock* needs tweaking. (The reason that this is not crucial to our current concerns is that we are not defending *Hitchcock* or any specific account of actual causation but merely using *Hitchcock* as a defeasible guide.)

But a second problem—more pressing precisely because it might seem to demand a default/deviant distinction—is that *Bogus Antidote* is structurally isomorphic to the core of a not-so-impoverished early pre-emption model:

$Throw_1$=1 if the first (pre-empting) vandal throws, 0 otherwise
$Throw_2$=1 if the second (backup) vandal throws, 0 otherwise
Shatter=1 if the window shatters, 0 otherwise

Early Pre-emption
$\boldsymbol{S_{early}}$ = <{$Throw_1$}, {$Throw_2$, *Shatter*}, $\boldsymbol{R_{early}}$>, where $\boldsymbol{R_{early}}$ maps all variables to {0, 1}
$\boldsymbol{L_{early}}$ = <$\boldsymbol{S_{early}}$, {$Throw_2$←1-$Throw_1$), *Shatter*←max($Throw_1$, $Throw_2$)}>
$\boldsymbol{M_{early}}$ = <$\boldsymbol{L_{early}}$, {$Throw_1$=1}>

Graphically we get:



And the real problem is that $\boldsymbol{M_{early}}$ is isomorphic to $\boldsymbol{M_{anti}}$, when they seem causally distinct: the early pre-emptor is a cause but the guard's administering the antidote in *Bogus Antidote* is not.[22]

We break the isomorphism by insisting that *Bogus Antidote* is impoverished (like *Bogus Prevention*) because it fails to represent whether or not the antidote neutralizes the poison and thereby ignores essential causal structure. Our reasoning is as follows. As noted above, we are reluctant to count Bodyguard's act as a cause of Victim's survival because it initiated a threat to Victim's life (a threat that it later counteracted). A proper model of the situation must capture this essential aspect of the causal structure of the case by representing the fact that Bodyguard's act endangered Victim's life. At a minimum, such a model should contain a path from *Antidote* to *Survival* such that, freezing off-path variables at certain values, setting *Antidote* at 1 (rather than 0) results in Victim dying (rather than surviving). But *Bogus Antidote* doesn't satisfy this constraint. It contains no path from *Antidote* to *Survival* such that, holding fixed off-path variables at certain values, setting *Antidote* = 1 makes *Survival* take value 0.

The problem disappears once we move to a richer model that includes a variable for neutralization, as follows:

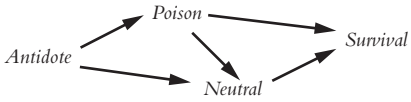
Bogus Antidote (Enriched)

$\boldsymbol{S_{anti+}}$ = <{*Antidote*}, {*Poison*, *Neutral*, *Survival*}, $\boldsymbol{R_{anti+}}$>, where $\boldsymbol{R_{anti+}}$ maps all variables to {0, 1}

$\boldsymbol{L_{anti+}}$ = <$\boldsymbol{S_{anti+}}$, {*Poison*←1-*Antidote*), *Neutral*←min(1-*Poison*, *Antidote*), *Survival*←max(Poison, Neutral)}>

$\boldsymbol{M_{anti+}}$ = <$\boldsymbol{L_{anti+}}$, {*Antidote*=1}>

Graphically one gets:



By including *Neutral*, *Bogus Antidote (Enriched)* lets us represent a contingency in which Victim's death depends on the administration of the antidote along a path. Consider the path from *Antidote* to *Survival* that doesn't go through *Neutral*. Setting *Neutral* = 0 (so that the antidote fails to neutralize the poison), we find that *Survival* takes value 0 just in case *Antidote* takes value 1. Our moral extends: the isomorphism is only arising due to the use of an impoverished and thus non-apt model. To use the language of Halpern and Hitchcock yet again, a lawyer seeking to defend

[22] Hall's (2007: 121–2) original isomorphism cases work by embedding a *Bogus Antidote* type short circuit in the one case, and an *Early Pre-emption* type structure in the other, into a larger and more complicated network. We are saying that Hall's (2007: 121) *Figure 9* is impoverished as a representation of short-circuit structure. Though the matter may be difficult to judge since Hall does not accompany his models with specific vignettes but characterizes the action abstractly.

Bodyguard's claim for saving Victim's life via *Bogus Antidote* would be subject to a devastating objection, of the form 'you've missed an essential part of the story, namely that had your client's antidote failed to neutralize the poison, Victim would have died precisely because of your client's actions'. The more pressing problem (for current purposes) is solved.[23]

Unfortunately even adding the neutralization part is not sufficient to save *Hitchcock* from still delivering the counterintuitive verdict that *Antidote*=1 causes *Survival*=1 in M_{anti+}. The reason is that on *Hitchcock,* the existence of counterfactual circumstances in which Victim would have died precisely as a result of Bodyguard's act has no relevance whatsoever to the causal status of the latter. The less pressing problem (for current purposes) remains. We suspect that an adequate theory of causation will have to be sensitive to the existence of such circumstances. But building such a theory is a very difficult matter, which we are not attempting here.[24]

Overall it seems to us that the wrong moral has been drawn from the existence of structurally isomorphic models of causally distinct cases. The right moral is to dump at least one of the two models invoked on the grounds that it fails to be apt. Indeed it seems to us that the following is a good heuristic: *When confronted with structurally isomorphic but causally distinct cases, suspect that at least one of the models is impoverished or otherwise non-apt*. This heuristic doesn't say which of the models is non-apt. That is the job of constraints like 7 and 8. Rather this heuristic functions

[23] Dmitri Gallow (personal communication) has raised a concern about how far our strategy extends, by asking us to reconsider the neuron diagrams from the original Hall cases, reimagined simply as possible worlds consisting entirely of the systems of neurons as diagrammed, subject to simple neuron-firing laws. We are not sure that we have a general answer to this sort of challenge, though we add that we are not sure that default-relativity is any help either in such cases, given that we have imagined away much of the actual world basis for regarding either state of the neuron as 'default'.

[24] Here is one strategy to consider, drawing on Sartorio (2005) and Weslake (forthcoming). Reconsider the intuitive reason why Bodyguard didn't cause the survival: namely that if neutralization hadn't occurred Victim would have died iff the antidote had been poured. This is equivalent to saying that had neutralization not occurred, Victim would have survived iff the antidote had not been administered. So the intuitive reason why Bodyguard's act is not a cause of the survival can also be expressed as follows: had Bodyguard not poured the antidote, then Victim's survival would still have depended on Bodyguard's omission (under a certain contingency). Perhaps, then, the lesson of the case is that we should require causes to make a difference in two senses. Not only should a cause make a difference to its effect; in addition, the alternative to the cause shouldn't make the very same difference to the effect as well. Here is a way of tweaking Hitchcock that formally implements this idea. First we make actual causation itself contrastive, replacing 'X=x causes Y=y' with 'X=x rather than X=x^* causes Y=y rather than Y=y^*'. We can continue to work with *Hitchcock* but now must amend 2c to read: '2c*. Had $\boldsymbol{Z}$ taken values $\boldsymbol{z}$ and X=x^*, then Y=y^*.' Secondly we impose a third and final condition (from Sartorio, independently useful for dealing with switching cases) that a cause X=x must make a difference in the following contrastive sense: '3. If the contrast X=x^* had happened instead of the cause X=x, then 'X=x^* rather than X=x would not meet conditions 1 and 2 for causing Y=y rather than Y=y^*.' The added condition 3 then rules out *Antidote*=1 rather than *Antidote*=0 as an actual cause of *Survival*=1 rather than *Survival*=0, on grounds that *Antidote*=0 rather than *Antidote*=1 would equally satisfy conditions 1 and 2 with respect to *Survival*=1 rather than *Survival*=0. We leave open whether or not such a strategy is ultimately viable. We only mean to illustrate the existence of possible tweaks to *Hitchcock* that satisfy the intuition that *Antidote*=1 (rather than *Antidote*=0) does not cause *Survival*=1 (rather than *Survival*=0), without invoking any default/deviant distinction.

as a useful 'warning signal' for the theorist that some non-apt model may be in use, which may trigger her to check both models more closely with her independently developed aptness constraints.

In general, we think that it is a mistake to reject a given system of representations on grounds that it *can conflate* distinct cases. (*Compare*: maps can conflate distinct territories if one zooms out far enough to miss all relevant internal structure. No one should reject a system of mapping for *that*. It is a poor case against Google maps that one can zoom out far enough to make Boston and New York both look like indistinguishable dots.) What it takes to show that a given system of representations is poor is showing that the system *cannot distinguish* distinct cases. (*Compare*: what it takes to show that a system of mapping is poor is that it cannot distinguish distinct territories, even at an appropriate level of resolution. What vindicates Google maps is that one can zoom in and recover the differences between Boston and New York.) We think that *Early Pre-emption (Impoverished)*, *Late Pre-emption (Impoverished)*, and *Bogus Prevention* are akin to crude causal maps that have zoomed out too far and missed relevant internal structure. So we think that the proper moral is not to reject the entire system of standard causal modelling, but instead to show how standard causal modelling can draw all the needed distinctions just by zooming in and using higher-resolution models.

3.4 Psychological Plausibility

So far we have argued that default-relativity introduces complicating and under-constrained unclarities, which are not ultimately useful for distinguishing the gardener from the queen or distinguishing the vandal from the guard. Of course we cannot rule out that default-relativity might be useful *for something*, but as of now we can only say that we see no such need yet established.[25]

But what about the idea that our judgements of actual causation are default (/norm) sensitive (§2.3)? Doesn't that itself provide some rationale for incorporating default-relativity into the account of actual causation, independent of the treatment of any specific cases? We think not. Indeed we think that a closer look at the relevant psychological approaches (e.g. availability theory) shows that incorporating default-relativity into the account of actual causation is psychologically *implausible*. Default-relativity is better understood as arising from a general and independent cognitive bias triggered by heuristics of causal cognition (and in many other domains), and not from our specific competence with the concept of actual causation. We conclude by explaining why.

[25] In fairness, Halpern and Hitchcock (forthcoming: §7.6) explore the idea of using a normality ranking over states to generate grades of causation, which they use to analyse further notions such as the legal notion of an *intervening cause*. We consider it premature to judge their proposal at this stage.

The crucial psychological data is that normative considerations influence causal judgements. For instance, consider *the pen vignette* from Knobe and Fraser (2008: 443; cf. Hitchcock and Knobe 2009: 594):

> The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take pens, but faculty members are supposed to buy their own.
>
> The administrative assistants typically do take the pens. Unfortunately, so do the faculty members. The receptionist repeatedly e-mails them reminders that only administrators are allowed to take the pens.
>
> On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later that day, the receptionist needs to take an important message...but she has a problem. There are no pens left on her desk.

Knobe and Fraser then observed that people tended to agree with the claim 'Professor Smith caused the problem' but tended to disagree with the claim 'The administrative assistant caused the problem'.[26] It seems very plausible that this pattern of agreement and disagreement is influenced by the normative difference in who is allowed to take the pens.

We accept the claim that normative differences (e.g. who is allowed to take pens) influence causal judgements. But there are multiple strategies for explaining how this influence works, with very different upshots for the psychological plausibility of default-relative causal models. Thus contrast:

> *Competence Strategy*: The influence of norms on causal judgements is to be explained by positing a role for norms in our concept of actual causation. The influence of norms on causal judgements is a matter of our competent use of this norm-laden concept.
>
> *Heuristics-and-Biases Strategy*: The influence of norms on causal judgements is not to be explained by positing a role for norms in our concept of actual causation. Rather the influence of norms on causal judgements is to be explained through (i) a norm-free concept of actual causation, (ii) cognitive heuristics for fast performance with this norm-free concept, plus (iii) general cognitive biases associated with these heuristics, in which norms come into play. The influence of norms on causal judgements is a matter of a general background cognitive bias influencing a heuristic for applying a norm-free concept.

Competence Strategy and *Heuristics-and-Biases Strategy* are not exhaustive, and there are also ways to combine elements of both. But for the sake of a tractable discussion we will focus on just these two strategies.

[26] Knobe and Fraser ran a between-subject design using a Likert scale from 1 (anchored at 'strongly disagree') to 7 ('strongly agree'). They observed a mean agreement for 'Professor Smith caused the problem' up at 5.2, and a mean agreement for 'The administrative assistant caused the problem' down at 2.8. (The difference was statistically significant.)

We are happy to allow that *Competence Strategy* would plausibly support default-relativity as being psychologically plausible. If competent use of the concept of actual causation itself involved consideration of default status, and drew on a conceptual distinction between default and deviant events, then default-relative models could boast the virtue of conforming to the contours of human thought. But by perfectly analogous reasoning, *Heuristics-and-Biases Strategy* would plausibly undermine default-relativity as being psychologically implausible. If our actual concept of causation is itself norm free, then default-relative models would display the vice of deviating from the contours of human thought. (In other words, to exactly the same extent as *Competence Strategy* would support default-relative models, *Heuristics-and-Biases Strategy* would support default-free models.)

So in order to argue that default-relativity is psychologically plausible (§2.4), one needs to do more than merely note normative influences on causal judgement. Given the availability of both *Competence Strategy* and *Heuristics-and-Biases Strategy*, which give opposite verdicts on the psychological plausibility of default-relativity, one needs to show that the normative influences on causal judgement are arising along the lines *Competence Strategy* suggests, rather than along the lines *Heuristics-and-Biases Strategy* suggests. In other words, one also needs to argue that the normative influence in question is arising *due to our competence with the concept of actual causation* rather than *due to background features of cognitive performance interacting with the use of this concept*. We do not think that this second step of the argument has even been considered in the literature so far, and moreover we doubt that it can be plausibly taken.

Indeed, we think that *Heuristics-and-Biases Strategy* is the clear best fit with existing psychological theorizing. In the now-classic availability framework of Tversky and Kahneman (1973) and Kahneman and Miller (1986), the task—as described by Kahneman and Miller (1986: 139)—is to give a constrained psychological account of 'the generation of alternatives to reality', in a way that explains (1986: 148) certain heuristics and biases of cognition, alongside the cognitive illusions these can generate. These heuristics and biases are thought to arise in general and systematically connected ways throughout domains such as frequency judgements, probability judgements, representativeness judgements, anecdotal reasoning, causal judgements, and counterfactual reasoning. (This is why availability theory is generally interesting in ways that go well beyond the psychology of causal judgement.)

In the availability framework one posits specific heuristics for a given domain. For instance with probability judgements, Tversky and Kahneman (1973: 207) posit a *representativeness* heuristic, on which an event is 'judged probable to the extent that it represents the essential features of its parent population or generating process'. These heuristics are domain-specific 'cheap and dirty' tricks that we use to get a decent answer fast. Such domain-specific heuristics then explain certain cognitive errors, in at least two ways. First, sometimes the heuristic is itself a source of error.

The probability of an event can differ from its representativeness. Secondly, the heuristic can introduce further sources of error. For instance, we are generally prone to mis-rating representativeness if we have just seen a dramatic image. As Tversky and Kahneman (1973: 230) comment: 'Many readers must have experienced the temporary rise in the subjective probability of an accident after seeing a car overturned by the side of the road.'

It may be useful to contrast now two ways of understanding availability effects on probability judgements. One strategy—which no one would endorse—would involve trying to incorporate the notions of default and deviant into the probability calculus itself:

> *Competence Strategy for Probability*: The influence of norms on probability judgements is to be explained by positing a role for norms in our concept of probability. The influence of norms on probability judgements is a matter of our competent use of this norm-laden concept.

The alternative, which is absolutely orthodox, involves keeping the probability calculus pure, but telling a psychological story involving domain-specific heuristics and general background cognitive biases:

> *Heuristics-and-Biases Strategy for Probability*: The influence of norms on probability judgements is not to be explained by positing a role for norms in our concept of probability. Rather the influence of norms on probability judgements is to be explained through (i) a norm-free concept of probability, (ii) cognitive heuristics for fast performance with this norm-free concept, plus (iii) general cognitive biases associated with these heuristics, in which norms come into play. The influence of norms on probability judgements is a matter of a general background cognitive bias influencing a heuristic for applying a norm-free concept.

(*Competence Strategy for Probability* and *Heuristics-and-Biases Strategy for Probability* are exactly the same as *Competence Strategy* and *Heuristics-and-Biases Strategy*, just with 'probability' substituted in for '(actual) causation' throughout.) It seems clear to us that one should keep the probability/causal calculus pure, and instead explain the influence of norms on judgements through general background cognitive biases.

Let us elaborate one way in which *Heuristics-and-Biases Strategy* might be developed in the domain of causal judgements. (This is hardly the only way to proceed, but represents a way we consider especially promising.) To begin with, it might be that the concept of actual causation is *contrastive*, with the structure 'c rather than c^* causes e rather than e^*' (Hitchcock 1996; Woodward 2003; Maslen 2004; Schaffer 2005; Menzies 2007; Northcott 2008, *inter alia*). On this approach, the concept of actual causation makes reference not just to the roles of *cause* and *effect*, but also to the further roles of *causal contrast* (for a specific alternative to the cause) and *effectual*

contrast (for a specific alternative to the effect). There is nothing normative in this contrastive approach.[27]

Secondly, it might be that we tend to employ a counterfactual heuristic for contrastive causal judgements, along the lines of: '*c* rather than *c** causes *e* rather than *e** iff: if *c** would have occurred then *e** would have occurred' (Schaffer 2005: 329).

Thirdly, there is strong independent evidence that normative considerations influence how easily people are able to access various alternatives for consideration in counterfactual reasoning. 'Deviant' events tend to leap out as especially salient to people and tend to trigger thoughts of the more normal alternative, while 'default' events tend to duck out of view and not trigger thoughts about alternatives at all.[28] Normative differences would then influence causal judgement, but not through competence with the concept of actual causation, but rather through an independent background feature of cognitive performance, namely the general *availability* of alternatives in cognition.

Putting this together, with *Heuristics-and-Biases Strategy for Probability* we saw a three-part account:

Concept: Probability (norm-free)
Heuristic: Representativeness
Bias: Normative considerations can influence the cognitive availability of counterfactual alternatives, and thereby influence our judgements of what is representative

So in the domain of causal judgements, as an elaboration of *Heuristics-and-Biases Strategy*, one might consider something like:

Concept: Causation (contrastive, norm-free)
Heuristic: Counterfactual judgement, if *c** had occurred, would *e** have occurred?
Bias: Normative considerations can influence the cognitive availability of counterfactual alternatives, and thereby influence counterfactual judgements

Indeed, since there is strong independent evidence (mentioned above) that normative considerations influence how easily people are able to access various alternatives for consideration in counterfactual reasoning, and since it is deeply plausible that we

[27] One could of course opt to add in something normative. For instance one could add in a requirement that the effectual contrast be a default. (There is no incoherence in adding default-relativity into a contrastive approach). But the contrastive approach we have in mind for illustrating this second strategy does not add in anything normative.

[28] A classic illustration of this phenomenon are the 'if only...' studies from Kahneman and Tversky (1982). Participants are given a vignette in which Jones is killed in a car crash with a 'drug-crazed teen'. In one version of the vignette Jones leaves the office at the normal time but follows an unusual route, while in the other version of the vignette Jones leaves the office at an unusual time but follows the usual route. Respondents were asked to imagine how the Jones family would continue an 'if only...' thought. Over 80 per cent of respondents who mentioned time/route followed the 'if only the unusual had been usual' pattern, leading Kahneman and Miller (1986: 143) to summarize: 'Exceptions tend to evoke contrasting normal alternatives, but not vice versa.' See McCloy and Byrne 2000 and Schaffer and Knobe 2012, 685–6 for further discussion.

use counterfactual heuristics in causal judgement, our basic strategy is entirely conservative.

To illustrate how this form of *Heuristics-and-Biases Strategy* comes together, consider Knobe and Fraser's pen vignette again. When asked whether Professor Smith caused the problem—which is not yet a causal claim about events—the contrastive view of causation has it that subjects need to consider a salient event involving Professor Smith, and a salient contrast event, as well as a salient event involving the problem (presumably just the occurrence of the problem) and a salient contrast to that (presumably things running smoothly). They then can be expected to use the heuristic: 'if the salient contrast event to the salient event involving Professor Smith (/the assistant) had occurred, would things have then run smoothly?' The explanation for the difference in judgements would be a cognitive bias explanation, exploiting differences in the cognitive availability of the salient events as well as their salient contrasts, for the professor as opposed to the assistant. Since 'deviant' events tend to leap out as especially salient to people and tend to trigger thoughts of the more normal alternative, this strategy predicts that the deviant actual event of Professor Smith's taking the pen will tend to leap to mind, and will tend to trigger thoughts of the default alternative of Professor Smith's not taking any pen; while by comparison the default actual event of the assistant's taking the pen will not tend to leap to mind so readily, and will not so readily tend to trigger thoughts of the deviant alternative of the assistant's not taking any pen.

But leaving aside any specific forms of *Heuristics-and-Biases Strategy* (of which there are many), the rationale for pursuing some form of *Heuristics-and-Biases Strategy* is not just to keep the probability/causal calculus pure, and not just to fit the structure of existing psychological theorizing, but to best explain the *generality and systematicity* of availability effects. On the rival *Competence Strategy*, the effect of availability (/the normative influence) is explained by positing specific features of the concept of actual causation. This means that, in order to account for the generality of availability effects, a theorist extending the first strategy would presumably wind up positing lots of individual concepts (the concept of frequency, the concept of probability, the concept of representativeness, etc.) that just so happened to each make reference to norms, and just so happened to each do so in similar ways, so as to generate similar availability effects. From the perspective of this sort approach it just looks like a *pure accident* that these many concepts each just so happen to make reference to norms in similar ways. A theorist invoking the second strategy instead has *a ready explanation for the generality and systematicity of normative influences*. After all, she is positing a single background feature of cognitive performance playing a role in all of these domains.[29]

[29] Indeed we note that most of the causal modellers drawn to the default-relative thesis *Menzies* have thought of it as a conservative extension of modelling techniques (§2.4) precisely on grounds that it only

In summary: The claim that causal models must distinguish default from deviant events—as championed by Menzies and encoded in the thesis we dub *Menzies*—has been said to provide a conservative and psychologically plausible extension of standard causal modelling, in ways that solve multiple problems. We have argued instead that this thesis generates complicating and under-constrained unclarities, while failing to solve the problems it has been claimed to solve, and while not fitting the most psychologically plausible accounts of how norms influence cognition generally. We remain open to the idea that causal models must distinguish default from deviant events, but we are far from convinced. We recommend that theorists first pay more attention to what counts as an apt causal model, as well as to accounts of how norms influence cognition generally, before adding more widgets into causal models.[30]

References

Beebee, H. 2004. 'Causation and Nothingness', in *Causation and Counterfactuals*, ed. J. Collins, N. Hall, and L. A. Paul. Cambridge, MA: MIT Press, 291–308.

Briggs, R. 2012. 'Interventionist Counterfactuals', *Philosophical Studies*, 160: 139–66.

Byrne, R. 2011. 'Counterfactual and Causal Thoughts about Exceptional Events', in *Understanding Counterfactuals, Understanding Causation*, ed. C. Hoerl, T. McCormack, and S. R. Beck. Oxford: Oxford University Press, 208–29.

Diaconis, P., Holmes, S., and Montgomery, R. 2007. 'Dynamical Bias in the Coin Toss', *SIAM Review*, 49: 211–35.

Hall, N. 2000. 'Causation and the Price of Transitivity', *Journal of Philosophy*, 97: 198–222.

Hall, N. 2007. 'Structural Equations and Causation', *Philosophical Studies*, 132: 109–36.

Halpern, J. 2000. 'Axiomatizing Causal Reasoning', *Journal of Artificial Intelligence Research*, 12: 317–37.

Halpern, J. 2008. 'Defaults and Normality in Causal Structures', in *Principles of Knowledge Representation and Reasoning: Proceedings of the Eleventh International Congress*, ed. G. Brewka and J. Lang. Menlo Park, CA: AAAI Press, 198–208.

Halpern, J., and Hitchcock, C. 2010. 'Actual Causation and the Art of Modeling', in *Heuristics, Probability, and Causality: A Tribute to Judea Pearl*, ed. R. Dechter, H. Geffner, and J. Halpern. London: College Publications, 383–406.

impacts treatments of actual causation and not the general norm-free background picture of causal structure. For instance, Halpern and Hitchcock (forthcoming: §5). fend off subjectivity complaints by saying: 'While our account of actual causation incorporates all of these elements [e.g., subjectivity and value-ladenness], actual causation is the wrong place to look for objectivity. Causal structure, as represented in the equations of a causal model, is objective.' The problem for these theorists is that there looks to be psychological evidence of normative influence on both judgements of actual causation and judgements of causal structure (since the latter are just counterfactual judgements). We find it hard to see a principled position that cares about psychological plausibility but yet divides 'subjective' actual causality from 'objective' counterfactuals.

[30] Thanks to Dmitri Gallow, Joseph Halpern, Christopher Read Hitchcock, Joshua Knobe, Jonathan Livengood, Robert Northcott, L. A. Paul, Melissa Renee Schumacher, and the MIT Work-In-Progress Group.

Halpern, J., and Hitchcock, C. Forthcoming. 'Graded Causation and Defaults', *British Journal for the Philosophy of Science*.

Halpern, J., and Pearl, J. 2005. 'Causes and Explanations: A Structural-Model Approach. Part I: Causes', *British Journal for the Philosophy of Science*, 56: 843–87.

Hart, H. L. A., and Honoré, A. M. 1985. *Causation in the Law*. Oxford: Oxford University Press.

Hiddleston, E. 2005. 'Causal Powers', *British Journal for the Philosophy of Science*, 56: 27–59.

Hitchcock, C. 1996. 'The Role of Contrast in Causal and Explanatory Claims', *Synthese*, 107: 95–419.

Hitchcock, C. 2001. 'The Intransitivity of Causation Revealed in Equations and Graphs', *The Journal of Philosophy*, 98: 273–99.

Hitchcock, C. 2007. 'Prevention, Preemption, and the Principle of Sufficient Reason', *The Philosophical Review*, 116: 495–532.

Hitchcock, C., and Knobe, J. 2009. 'Cause and Norm', *Journal of Philosophy*, 106: 587–612.

Kahneman, D., and Miller, D. 1986. 'Norm Theory: Comparing Reality to its Alternatives', *Psychological Review*, 93: 136–53.

Kahneman, D., and Tversky, A. 1982. 'The Simulation Heuristic', in *Judgement Under Uncertainty: Heuristics and Biases*, ed. D. Kahneman, P. Slovic, and A. Tversky. New York: Cambridge University Press, 201–11.

Knobe, J., and Fraser, B. 2008. 'Causal Judgement and Moral Judgement: Two Experiments', in *Moral Psychology, vol. 2: The Cognitive Science of Morality: Intuition and Diversity*, ed. W. Sinnott-Armstrong. Cambridge, MA: MIT Press, 441–8.

Lewis, D. 1986a. 'Causation', in his *Philosophical Papers vol. 2*. Oxford: Oxford University Press, 159–213.

Lewis, D. 1986b. 'Events', in his *Philosophical Papers vol. 2*. Oxford: Oxford University Press, 241–69.

Livengood, J. 2013. 'Actual Causation and Simple Voting Scenarios', *Noûs*, 47: 316–45.

McCloy, R., and Byrne, R. 2000. 'Counterfactual Thinking about Controllable Events', *Memory & Cognition*, 28: 1071–8.

McDermott, M. 1995. 'Redundant Causation', *British Journal for the Philosophy of Science*, 40: 523–44.

McGrath, S. 2005. 'Causation by Omission', *Philosophical Studies*, 123: 125–48.

Mackie, J. L. 1974. *The Cement of the Universe*. Oxford: Oxford University Press.

Maslen, C. 2004. 'Causes, Contrasts, and the Nontransitivity of Causation', in *Causation and Counterfactuals*, ed. J. Collins, N. Hall, and L. A. Paul. Cambridge, MA: MIT Press, 341–57.

Maudlin, T. 2004. 'Causation, Counterfactuals, and the Third Factor', in *Causation and Counterfactuals*, ed. J. Collins, N. Hall, and L. A. Paul. Cambridge, MA: MIT Press, 419–43.

Menzies, P. 1989. 'Probabilistic Causation and Causal Processes: A Critique of Lewis', *Philosophy of Science*, 56: 642–63.

Menzies, P. 2004. 'Difference-Making in Context', in *Causation and Counterfactuals*, ed. J. Collins, N. Hall, and L. A. Paul. Cambridge, MA: MIT Press, 139–80.

Menzies, P. 2007. 'Causation in Context', in *Causation, Physics and the Constitution of Reality: Russell's Republic Revisited*, ed. H. Price and R. Corry. Oxford: Oxford University Press, 191–223.

Menzies, P. 2009. 'Platitudes and Counterexamples', in *The Oxford Handbook of Causation*, ed. H. Beebee, C. Hitchcock, and P. Menzies. Oxford: Oxford University Press, 341–67.

Menzies, P. 2011. 'The Role of Counterfactual Dependence in Causal Judgments', in *Understanding Counterfactuals, Understanding Causation*, ed. C. Hoerl, T. McCormack, and S. R. Beck. Oxford: Oxford University Press, 186–207.

Mill, J. S. 1950. *A System of Logic*. London: Macmillan Publishers.

Northcott, R. 2008. 'Causation and Contrast Classes', *Philosophical Studies*, 139: 111–23.

Paul, L. A., and Hall, N. 2013. *Causation: A User's Guide*. Oxford: Oxford University Press.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Sartorio, C. 2005. 'Causes as Difference-Makers', *Philosophical Studies*, 123: 71–96.

Sartorio, C. 2010. 'The Prince of Wales Problem for Counterfactual Theories of Causation', in *New Waves in Metaphysics*, ed. A. Hazlett. London: Palgrave Macmillan, 259–76.

Schaffer, J. 2004. 'Causes Need Not be Physically Connected to their Effects: The Case for Negative Causation', in *Contemporary Debates in Philosophy of Science,* ed. C. Hitchcock. Oxford: Basil Blackwell, 197–216.

Schaffer, J. 2005. 'Contrastive Causation', *The Philosophical Review*, 114: 327–58.

Schaffer, J. 2010. 'Contrastive Causation in the Law', *Legal Theory*, 16: 259–97.

Schaffer, J. 2012. 'Disconnection and Responsibility', *Legal Theory*, 18: 399–435.

Schaffer, J., and Knobe, J. 2012. 'Contrastive Knowledge Surveyed', *Noûs*, 46: 675–708.

Shulz, K. 2011. '"If you'd wiggled A, then B would've changed": Causality and Counterfactual Conditionals', *Synthese*, 179: 239–51.

Spirtes, P., Glymour, C., and Scheines, R. 1993. *Causation, Prediction, and Search*. New York: Springer-Verlag.

Tversky, A., and Kahneman, D. 1973. 'Availability: A Heuristic for Judging Frequency and Probability', *Cognitive Psychology*, 5: 207–32.

Weslake, B. forthcoming. 'A Partial Theory of Actual Causation', *British Journal for the Philosophy of Science*.

Woodward, J. 2003. *Making Things Happen*: *A Theory of Causal Explanation*. New York: Oxford University Press.