Actual Causation by Probabilistic Active Paths

Charles R. Twardy and Kevin B. Korb[†‡]

Recently Halpern and Pearl and Hitchcock have presented influential accounts of actual (token) causation using Bayesian networks. These accounts have been deterministic. Here we present a probabilistic extension to these active path analyses of actual causation. The extension uses "soft" interventions to set distributions rather than just single values. The resulting account can handle at least as wide a range of examples as the original accounts, without assuming determinism.

1. Introduction. Given a probabilistic causal model that describes the causal structure for many possible situations, how does one assign credit or blame for what actually happens? Causal models have a natural semantics: variable C is a cause if and only if (iff) there is some context in which a change in the value of C would make a difference to E (see, e.g., Woodward 2005). But in actual causation, we do not care that "there is some context." We want an answer for this actual context. Certainly smoking can kill, but did Jones's smoking kill him?

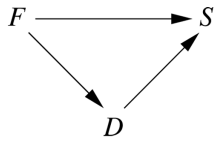
In Handfield et al. (2008), we argued for integrating causal models with process-theory semantics. Here we attempt to push stochastic causal models as far as they can go alone. Although finally incomplete, the account can surpass the limits declared by Hiddleston and others. We demonstrate the model on several problematic examples from the literature, including

†To contact the authors, please write to: Charles R. Twardy, C[4]I Center, George Mason University, 4400 University Dr., MS 4B5, Fairfax, VA 22030; e-mail: ctwardy@c4i.gmu.edu; Kevin B. Korb, School of Info Tech, Monash University, Clayton, VIC 3800, Australia; e-mail: kbkorb@gmail.com.

‡This work was funded in part by a Monash Arts/IT grant. We are grateful for comments by Toby Handfield, Graham Oppy, Lucas Hope, Chris Hitchcock, an anonymous reviewer, and participants in the Monash 2004 causation seminar as well as PSA 2010 in Montreal.

Philosophy of Science, 78 (December 2011) pp. 900–913. 0031-8248/2011/7805-0016$10.00
Copyright 2011 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/662957 Published online by Cambridge University Press



$$F = 1$$
$$D = F$$
$$S = \neg F \vee D$$

Figure 1. Boulder example. D = ducks; F = falls; S = survives.

cases of overdetermination and "late cutting" (where a slower potential cause is rendered otiose by a faster-acting cause).

This article extends "active path" accounts such as those of Hitchcock (2001) and Halpern and Pearl (2005a, 2005b). The key idea is that in a given context, some paths become inactive. An actual cause must be on an active path. But earlier accounts assume deterministic networks, which is unnecessary and contrary to the spirit of probabilistic causal models. Determinism also masks some subtleties, much as using binary variables can mask intransitivity.

2. Deterministic Active Paths. Following Hiddleston (2005), we use H1 for Hitchcock's "strongly active path," H2 for Hitchcock's (2001) "weakly active path" with redundancy ranges, and H3 for Halpern and Pearl (2005a, 2005b). We will present a reconstruction that blends H2 and H3.

2.1. Strongly Active Paths (H1). A boulder falls (F) toward a hiker (Hitchcock 2001). The hiker sees the boulder in time, and ducks (D), thereby surviving (S). Each variable is binary. We represent the causal variables as nodes in a graph and the causal links as arrows between them. We specify the exact dependencies with a series of simple structural equations, as in figure 1. The first equation gives an initial condition: in the actual world, the boulder falls. The second says that the hiker ducks if the boulder falls. The third says the hiker survives if he ducks or if the boulder does not fall. The causal model is fully specified by the equations and the possible values for each variable.

Causal models specify the counterfactuals over the variables in the model, in the same way a circuit diagram specifies the effect of any alteration or failure of its components (Pearl 2000). Counterfactuals do not backtrack. If they did, ducking would make no difference to survival. But counterfactuals are assessed by *interventions*, not observations. We do not observe that D is 1. Instead, we replace the structural equation $D = F$ with $D = 1$ and reevaluate. The boulder falls ($F = 1$), and $S = 1$ when $D = 1$, and $S = 0$ when $D = 0$, so ducking makes a difference. Therefore, we want to say that ducking caused survival. Conversely, we do not want

https://doi.org/10.1086/662957 Published online by Cambridge University Press

to say that the boulder falling caused survival. The difference is that only ducking has an active path to survival.

We now define active path actual causation (APAC). A *path* is a traversal from C to E along the arrows, in the direction of the arrows. Then:

> **Definition 1 (APAC).** $C = c$ was an actual cause of $E = e$ iff there was an active path ϕ from C to E (and actually $C = c$ and $E = e$).

Intuitively, a path is active if it makes a difference. A path ϕ is strongly active if c makes a difference to e when all other paths between C and E are blocked by setting variables along those paths to their actual values. "Making a difference" has the usual counterfactual meaning: had C been different, E would have been different, for at least some $c' \neq c$.

In Boulder, neither of the paths from F to S is strongly active. First, we hold D fixed at its actual value ($D = 1$) blocking $F \rightarrow D \rightarrow S$ and leaving only $F \rightarrow S$. But since $D = 1$, F makes no difference: given that the hiker ducked, the boulder has no effect. The path $F \rightarrow S$ is not strongly active.

To evaluate $F \rightarrow D \rightarrow S$ for strong activity, we would have to block $F \rightarrow S$, which we cannot, as there is no intermediate variable. With nothing to block, we are back to the original case where F makes no difference to S because the paths cancel. We happily conclude that F did not cause S. "Intuitively, the falling boulder does not save Hiker's life because without it, Hiker's life would not have been endangered in the first place" (Hitchcock 2001, 295).

When we are considering a particular path, all ancestors of E not on it constitute the *background*. A straightforward (if inefficient) attempt to block all but one path is to intervene to set the whole background—all ancestors of E except those along our path ϕ. Call this background β.

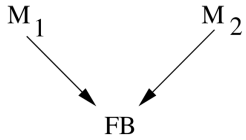
> **Definition 2 (Background β).** Given a target variable E and path ϕ ending with E, the background β is all the ancestors of E except for those in ϕ.

> **Definition 3 (β-actual model).** Given target variable E, the β-actual model replaces all the equations for β with their actual values **b**. The replacements are by intervention.

We can now give a formal definition of strong activity. Let $I_{C=c}$, or just I_c, denote that we intervene to set $C = c$; similarly for $I_{c'}$. (We shall also use "set" as an English equivalent for the I notation.) We assume throughout that C is the beginning of the path ϕ and E its end. Then:

> **Definition 4 (Strongly active path).** A path ϕ is strongly active iff I_c

https://doi.org/10.1086/662957 Published online by Cambridge University Press



$$\begin{aligned} M_1 &= 1 \\ M_2 &= 1 \\ FB &= M_1 \vee M_2 \end{aligned}$$

Figure 2. Matches example. FB = forest burns; M = match.

(vs. $I_{c'}$) makes a difference to E, in the β-actual model, for some c and c'.

If we allow only strongly active paths, then APAC yields Hitchcock's first active path account, which Hiddleston (2005) dubs H1. Unfortunately, H1 handles neither symmetric overdetermination nor late cutting. We review *redundancy ranges* to handle overdetermination and then *resetting ranges* to avoid late cutting. Afterward, we present our probabilistic reformulation.

2.2. Redundancy Ranges. Symmetric overdetermination happens, given multiple sufficient causes. One such case is Matches. Two lit matches (M_1 and M_2) are dropped in the forest. Either would suffice to burn it down, and the forest burns down (FB; fig. 2).

The reader can verify that neither match has a strongly active path. However, intuitively both matches contributed to the fire, although neither alone was necessary. The causal structure allows us to reveal the hidden dependency.

Note, we can let M_2 be 0 without affecting the result—the forest still burns down. Hitchcock calls the set of such "redundant" settings of background variables the redundancy range.

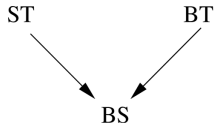
Definition 5 (Redundancy range). Given a particular path ϕ and background β, $\mathbf{b}'$ lies in β's redundancy range for E iff $I_{\mathbf{b}'}$ leaves E unchanged.

The redundancy range is the set of all possible ways of fixing the background variables so that they make no difference to E. Of course, the actual values of β are also in the redundancy range.

Definition 6 (β model). The $\mathbf{b}'$ model replaces all $B \in \beta$ with values $\mathbf{b}'$ in their redundancy range, via $I_{\mathbf{b}'}$.

In Matches, by setting $M_2 = 0$, we reveal a hidden dependency: now M_1 makes a difference. We define an active path (simpliciter) as a path

https://doi.org/10.1086/662957 Published online by Cambridge University Press



$$ST = 1$$
$$BT = 1$$
$$BS = ST \vee BT$$

Figure 3. Bottle example. BS = bottle shatters; BT = Billy throws; ST = Suzy throws.

that becomes strongly active when we set background variables to some values in their redundancy range (H2).

> **Definition 7 (Active path)**. A path ϕ is active iff I_c (vs. $I_{c'}$) makes a difference to E in some $\mathbf{b}'$ model. More precisely, $I_{c,\mathbf{b}'}$ gives e and $I_{c',\mathbf{b}'}$ gives $e' \neq e$.

Both paths in Matches are active, so both matches are causes according to definition 1. But we cannot yet handle late cutting.

2.3. Resetting Ranges (∼H3). Late cutting is a kind of preemptive overdetermination. It occurs when one of the potential causes preempts the other, precisely by completing first. For example, Bottle:

> Suzy and Billy throw rocks at a bottle. Suzy's rock arrives first, breaking the bottle before Billy's rock, which would also have broken it.

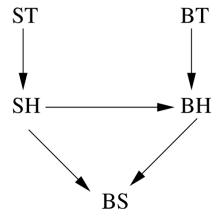
We want to say that Suzy's throw (ST) caused the bottle to break (BS) but that Billy's throw (BT) did not. First, note that the obvious model (fig. 3) does not work. This is the same model as Matches, so it gives the same answer: both throws caused the bottle to smash. But our intuitions strongly favor Suzy's throw over Billy's: the preemption matters.

Halpern and Pearl (2005a) argue that this model misrepresents the case since the model is symmetric and Bottle is not. They argue that a proper model should show the key asymmetry: when Suzy's throw hits (SH), it prevents Billy's from hitting (BH). Therefore, Halpern and Pearl suggest the model in figure 4.

This seems a good model, and it works for Suzy, whose path is strongly active. Unfortunately, Billy's path (BT → BH → BS) is also active. Let $\mathbf{b}' = \{\text{ST} = 0; \text{SH} = 0\}$: $\mathbf{b}'$ is in the redundancy range, and under $I_{\mathbf{b}'}$, BT makes a difference.

Hitchcock's H2 solves that problem by requiring that $\mathbf{b}'$ not change anything on the path ϕ, whereas $\mathbf{b}'$ implies a change to BH. But Halpern and Pearl showed that this strategy was vulnerable to a voting machine example (Twardy and Korb 2011). Their solution is to allow resetting of

https://doi.org/10.1086/662957 Published online by Cambridge University Press



$$
\begin{aligned}
ST &= 1 \\
BT &= 1 \\
SH &= ST \\
BH &= BT \wedge \neg SH \\
BS &= SH \vee BH
\end{aligned}
$$

Figure 4. Modified bottle example. BH = Billy hits; BS = bottle shatters; BT = Billy throws; SH = Suzy hits; ST = Suzy throws.

variables along ϕ. We now present a somewhat simplified version of resetting variables.

Definition 8 (H3AP). ϕ is active iff:

A. When we set c' and $\mathbf{b}'$ as in definition 7, we get $e' \neq e$.
B. When we set c and $\mathbf{b}'$, we get e, even if we reset any of the other variables in ϕ to their actual values.

Note that resetting is only done for $C = c$; otherwise, no cause could make a difference.

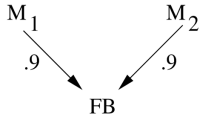
Condition A amounts to saying that there is a background state ($\mathbf{b}'$) where c can make a difference to e. Condition B ensures that the background $\mathbf{b}'$ alone is insufficient to force $E = e'$: it means that $\mathbf{b}'$ is in the redundancy range for e. The "even if" clause about resetting variables in ϕ to their actual values is only necessary in cases of late cutting.

In the extended bottle-smashing model, Suzy's throw ST = 1 remains an actual cause of BS = 1 using a redundancy range to reveal the dependency. Choosing {BT, BH} as our background, and setting BH = 0, we see that ST makes a difference. Resetting has no effect because, actually, SH = 1.

Billy's throw would similarly make a difference, save for the new "resetting" clause. Let the background be {ST, SH}. Setting SH = 0, Billy's throw makes a difference. However, we can reset BH to its original value when BT = 1. Of course, then Billy's throw will no longer make a difference, and BS = 0, in violation of clause B of definition H3AP. Therefore, this path is not an active path, and BT is not an actual cause. We turn now to probabilistic extensions.

3. Probabilistic Active Paths. The counterfactuals in H1–H3 are of the form, "if not ($C = c$), then not ($E = e$)." That will not work for probabilistic causation. In Matches, if each match acts independently with probability p of burning down the forest, then both matches will have a

https://doi.org/10.1086/662957 Published online by Cambridge University Press



$$\begin{aligned} M_1 &= 1 \\ M_2 &= 1 \\ \Pr(FB) &= 1-(1-s)^{M_1}(1-t)^{M_2} = 0.99 \end{aligned}$$

Figure 5. Probabilistic matches example. FB = forest burns; M = match.

greater probability of burning down the forest, but it still might not burn down. Conversely, the model could allow the forest to burn down with neither match lit: active paths need to be redefined in terms of changing probabilities rather than changing states. Except for noting that by "cause" we will henceforth mean "causally relevant," we retain APAC. (We can distinguish promoters from preventers by examining whether the probability went up or down.)

We will use two measures of probabilistic difference making. The first measure, Δ_1, is the difference caused by c in some background $\mathbf{b}'$.

$$\Delta_1 = \Pr(e|I_{c,\mathbf{b}'}) - \Pr(e|I_{c',\mathbf{b}'}).$$

Intuitively, a path is active when $\Delta_1 \neq 0$. However, with finite data to estimate probabilities and limited computing power, we cannot be so stringent, so we introduce a tolerance parameter. Let ε denote some threshold below which we consider changes in probability to be insignificant. Then, a path is active when $|\Delta_1| > \varepsilon$. (If $\Delta_1 < -\varepsilon$, then the path is an active *preventer*.)

To be strongly active, we must have $|\Delta_1| > \varepsilon$ in the β-actual model, when $\mathbf{b}' = \mathbf{b}$. But in general, $\mathbf{b}'$ can be any member of the redundancy range, which are instantiations of β whose effect on $\Pr(e)$ is less than ε.

We define Δ_2 to be the effect of wiggling $\mathbf{b}'$ versus $\mathbf{b}$ in "background" c:

$$\Delta_2 = \Pr(e|I_{\mathbf{b},c}) - \Pr(e|I_{\mathbf{b}',c}).$$

Instantiation $\mathbf{b}'$ lies in β's redundancy range iff $|\Delta_2| \leq \varepsilon$. This already is sufficient for handling probabilistic matches (see fig. 5).

Let s and t be the probabilities that M_1 and M_2, respectively, acting alone would burn down the forest (i.e., in this case $s = t = 0.9$). The model combines these probabilities independently in determining the probability of FB (i.e., there is no "synergy" between the two causes). The model also assumes the forest cannot burn down spontaneously (or, indeed, by any other cause than the matches named); this is recorded in the third equation, where the exponents just switch the terms on and off. That equation expresses the idea that M_1 and M_2 act independently since their chances of failure simply multiply.

Variables M_1 and M_2 are only strongly active (i.e., active relative to the

https://doi.org/10.1086/662957 Published online by Cambridge University Press

TABLE 1. M_1 AS THE CANDIDATE CAUSE.

	$\mathbf{b}{:}M_2 = 1$		$\mathbf{b}'{:}M_2 = 0$	
ε Range	$\Delta_1 > \varepsilon$? ($\Delta_1 = .09$)	$\Delta_2 \leq \varepsilon$? ($\Delta_2 = 0$)	$\Delta_1 > \varepsilon$? ($\Delta_1 = .9$)	$\Delta_2 \leq \varepsilon$? ($\Delta_2 = .09$)
[.9, 1]	No	Yes	No	Yes
[.09, .9)	No	Yes	Yes	Yes
[0, .09)	Yes	Yes	Yes	No

other match being lit) for $\varepsilon \leq .09$. We can construct table 1, relative to M_1 as the candidate cause. There are two cases when M_1 is counted as a cause, depending on ε. If $\varepsilon < .09$, then M_1 is a cause because its path is active in the actual case ($\beta = \mathbf{b}{:}M_2 = 1$). If $.09 \leq \varepsilon < .9$, then M_1 is a cause because its path is active in the redundant case ($\beta = \mathbf{b}'{:}M_2 = 0$). If $\varepsilon \geq .9$, then M_1 is not a cause. The analysis is symmetrical for M_2. Therefore, if $\varepsilon < .9$, both matches cause the forest to burn down, but if $\varepsilon \geq .9$—a very large tolerance—then neither does.

We can make the model "feel" more deterministic by defining $\mathrm{FB} = \max(s, t)$. In that case, $\Pr(\mathrm{FB}) = .9$ so long as any match is lit. Clearly, neither is strongly active for any ε. Nevertheless, both are active simpliciter for any $\varepsilon < 0.9$. Setting $s = t = 1$ recovers the original deterministic Matches exactly, concluding that both matches are actual causes (so long as $\varepsilon \neq 1$); the same, of course, can be said of the original model.

3.1. Soft Interventions. The final step will be to introduce resetting ranges, dealing with late cutting. However, for assessing active paths probabilistically, we turn to "soft interventions." A soft intervention (Korb et al. 2004) fixes a variable to a distribution, instead of a single value. For example, fixing $\mathrm{SH} = \{0.7, 0.3\}$ (i.e., a 70% chance of Suzy hitting) is a soft intervention. The proper probabilistic analog of "setting X back to the value it would have under c" is "setting X back to the distribution it would have under c."

Soft interventions are easily dealt with formally. A soft intervention on X is implemented by adding a new parent $I \rightarrow X$, such that when I is true, $\Pr(X)$ takes the desired distribution, and when I is false, X has its previous distribution. How this might be done experimentally (outside of simulation) is less clear. Nevertheless, many actual interventions are soft rather than hard (Korb et al. 2004). In any case, we are not here concerned with implementation but with using soft interventions as a calculating tool.

The soft interventions of interest here freeze X at its current distribution. Because the distribution is unchanged, such interventions have no downstream effect. However, all back paths are severed: changes in parent values (other than I) have no effect on X; changes in X can only

https://doi.org/10.1086/662957 Published online by Cambridge University Press

then tell us about I and not its other parents or, through them, any cousin variables. This gives us exactly what we need: the ability to distinguish causes and effects by the asymmetries of intervention. In this article, we denote such distributional (soft) interventions with a vector symbol $\vec{c}$—or $I_{\vec{c}}$—denoting that we have simply replaced the distribution for C. (When we do not specify that new distribution, by default we shall mean the current distribution.)

3.2. Resetting Ranges and Probabilistic Active Paths. Let there be n variables along path ϕ, not counting C and E. There are 2^n possible combinations for clamping (or not). Let $\mathcal{R}$ denote the set of possible "resetting ranges," the power set listing all possible combinations of these variables. Then for $\mathbf{r} \in \mathcal{R}$, let $I_{\vec{\mathbf{r}}}$ denote the intervention resetting those variables in $\mathbf{r}$ to their current distribution. We can now define the probabilistic active paths (PAP) criterion:

Definition 9 (PAP). A path ϕ is active iff for some background $\mathbf{b}'$:

1. $|\Delta_1| > \varepsilon$. (Wiggling C makes a difference.)
2. $|\Delta_2| \leq \varepsilon$. (Wiggling $\mathbf{b}'$ alone does not make a difference.)
3. In $I_{c,\mathbf{b}'}$, $I_{\vec{\mathbf{r}}}$ leaves $|\Delta_2| \leq \varepsilon$, $\forall \mathbf{r} \in \mathcal{R}$.

Condition 3 requires that condition 2 holds for all combinations of variables in ϕ; that is, resetting them to their original distributions leaves $\mathbf{b}'$ in the redundancy range. As noted, we add our clamp $I_{\vec{\mathbf{r}}}$ to the interventions already done for Δ_2. A brute force search over the power set of resetting ranges will naturally be expensive, but we are not concerned with that here. For the small models treated in this literature, the problem is not serious.

4. Probabilistic Examples. Here we review the application of PAP to some examples from the literature. PAP deals well with deterministic examples (if H3 does) because PAP with degenerate distributions is H3. So here we treat only probabilistic variants of the examples.

4.1. Probabilistic Bottle. For a more realistic version of Bottle, we allow that Suzy and Billy sometimes miss. In particular, given knowledge only of ST and BT, we use table 2. We use the extended model of Halpern and Pearl to capture the timing asymmetry as before, yielding figure 6.

Let $\varepsilon = 0$, so the redundancy ranges are degenerate for both Suzy and Billy: we have to use the actual background BT = 1 or ST = 1. The network is shown in three states in figure 6. The original distribution is in figure 6*a*. As before, $\Delta_1 > \varepsilon$ and $|\Delta_2| < \varepsilon$ for both Suzy and Billy. Next,

https://doi.org/10.1086/662957 Published online by Cambridge University Press

TABLE 2. CONDITIONAL PROBABILITY.

ST	BT	Pr (BS = 1)
0	0	0
0	1	.8
1	0	.7
1	1	.94

Note.—BS = bottle shatters; BT = Billy throws; ST = Suzy throws.

we see whether resetting the intermediate "hits" variables violates condition 3.

In figure 6*b*, we consider Suzy, resetting to SH = {0.7, 0.3}. This leaves Pr (BS) unchanged, so we conclude that Suzy Throws is an actual cause. In figure 6*c*, we consider Billy. Resetting to BH = {.24, .76}, Pr (BS) changes, so $\Delta_2 > \varepsilon$, and BT is not actually a cause on PAP. This happens because the intervention on BH blocks the backward flow of probability otherwise relating BS and BH through their common cause SH. Before resetting, the distribution BH = {.24, .76} arose because Billy could hit only when Suzy did not. Suzy missed 30% of the time, and 0.8 × 30% = 24%. Now, however, Billy never hits more than 24% of the time, so Pr (BS = 1) drops from 94% to 77%.

The original probabilistic Boulder can be resolved without resetting. However, if we add noise at *S*, so that ducking no longer guarantees survival, we find an influence for *F*, which we argue is appropriate. (For further analysis, see Twardy and Korb 2011, the online companion piece.)

4.2. The Trainee Assassin. The deterministic version is that a trainee assassin *T* shoots a victim *V*, but Supervisor will shoot if Trainee loses nerve. Being deterministic, Victim survives only when neither shot is fired. In the actual case, the trainee shoots, the supervisor does not, and the victim dies. As with Boulder, there is no marginal effect, but any path-based account works because the path $T \rightarrow V$ is strongly active. However, Hiddleston (2005) alleges that the analysis will not survive a probabilistic extension, which leads us to the stochastic assassin.

4.3. The Stochastic Assassin. In the probabilistic version, Trainee firing lowers the overall chance that Victim dies, by preventing the more accurate Supervisor from shooting. Once again, a path-based analysis of actual causation does fine here. Let $T \xrightarrow{.5} V$ and $S \xrightarrow{.9} V$. For completeness, $T \xrightarrow{.9} \neg S$.

Since Supervisor does not shoot, Trainee is strongly active with $\Delta_1 = 0.5$. Similarly, Supervisor's $S = 0$ (vs. $S = 1$) has $\Delta_1 = -0.45$ in the actual background where Trainee shoots, and so is a strongly active pre-

https://doi.org/10.1086/662957 Published online by Cambridge University Press

https://doi.org/10.1086/662957 Published online by Cambridge University Press

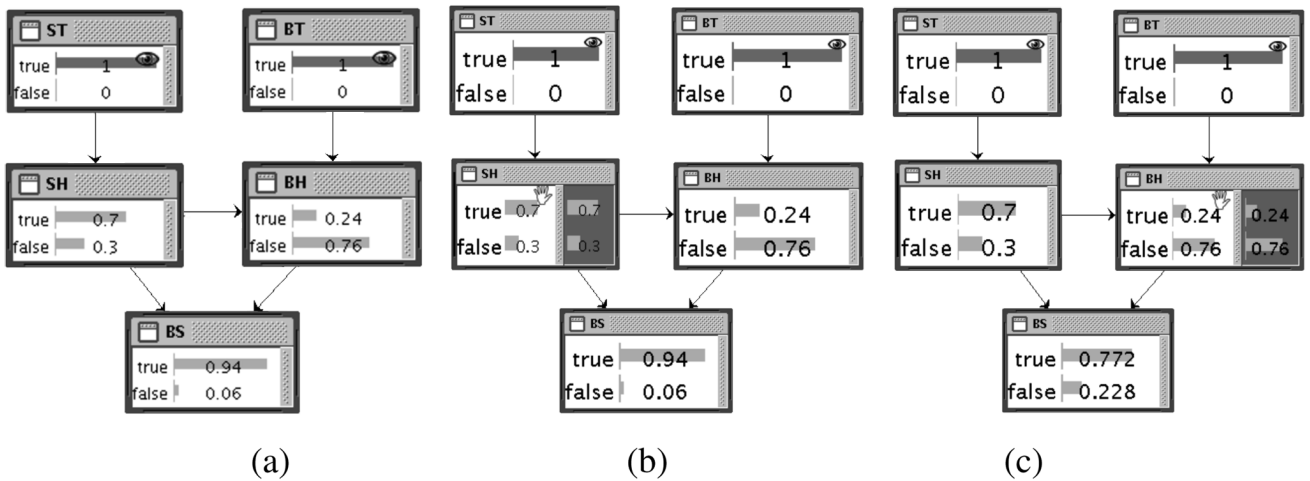
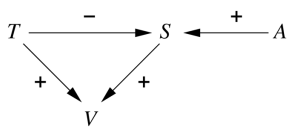


Figure 6. Probabilistic version of bottle with results of intervening to fix the probability distributions: *a*, original state; *b*, clamping SH has no effect on BS; *c*, clamping BH affects BS. BH = Billy hits; BS = bottle shatters; BT = Billy throws; SH = Suzy hits; ST = Suzy throws.



T	A	$\Pr(S = 1)$
0	0	.85
0	1	.95
1	0	.05
1	1	.15

Figure 7. Stochastic assassin example, with supervisor's conditional probability table. A = aiming; S = supervisor; T = trainee; V = victim.

venter (of death), having the same status as the falling boulder. That is, Supervisor is ineluctably causally active.

Hiddleston recommends adding an extra variable for Supervisor taking Aim (A). The supervisor is unlikely to fire without aiming and is always more likely to fire if she aims. Avoiding extremes, let Supervisor be 90% likely to fire if Trainee loses nerve and 10% likely even if Trainee also fires. Let those numbers average over Aiming and not-Aiming, with a 50 : 50 prior over each. Furthermore, let aiming improve chances 10% over not aiming in both cases. The model and conditional probability table are shown in figure 7.

As before, let T and S be independent causes of V (in a noisy-or model). Consider whether $A = 1$ is an actual cause of $V = 1$. In the actual background, $T = 1$, and $\Delta_1(A) = .914 - .905 = .009$. Aiming appears to be an actual cause, provided $\varepsilon < .009$. Clamping S at the distribution $\{.15, .85\}$ induced by $\{A, T\}$ has no effect on $\Pr(V)$. Therefore, $A = 1$ is on an active path and is a cause.

Hiddleston thinks this is wrong because in actual fact Supervisor did not shoot, so Aiming made no difference. We disagree. As our analysis has made clear, when considering the probabilistic effect of C on E, one must unfix the path between C and E. If we fixed S, we would cut the active path and get Hiddleston's answer, but in a probabilistic context that procedure implies that only direct causes are actual causes. Hiddleston is correct that such probabilistic accounts cannot work, but that is just a misconstrual of probabilistic accounts. Hiddleston's superfluous antidote presents more of a problem.

4.4. Superfluous Antidote. Fred injects himself with a harmless antidote ($A = 1$) in the false belief that he has been (or is about to be) poisoned (P). The model is simple: $P \rightarrow S \leftarrow A$, and $S = \neg P \vee A$. Fred survives when $S = 1$. In fact there is no poison ($P = 0$), so the antidote is irrelevant. However, because $A = 1$, the redundancy range includes $P = 1$. Were there poison, the antidote would make a difference; hence, the path $A \rightarrow S$ is active. As Hiddleston says, "That is the wrong answer. H2 [and

https://doi.org/10.1086/662957 Published online by Cambridge University Press

H3] counts ($A = 1$) an 'actual cause' merely because it *would have been* a cause if the circumstances had been different" (2005, 32).

This is a good counterexample because it targets redundancy ranges, an epicycle added to handle symmetric cases. H1, which lacks redundancy ranges, correctly says $A = 1$ is not a cause. But it also says $P = 0$ is not a cause, leaving $S = 1$ uncaused, as it does with Matches. Several responses might mitigate the damage:

- Pick one's poison, trading off cases like Matches for those like Superfluous Antidote. This is unsatisfactory since nothing is mitigated.
- Make the problem probabilistic. For many reasonable values of ε, $\Delta_1 < \varepsilon$ in the actual background $P = 0$, or $\Delta_2 > \varepsilon$ if we try to set $P = 1$. That solves the problem. However, the point of the example involves extreme probabilities, which this response ignores.
- Hiddleston recommends employing Cheng's account of causal power (Cheng 1997).
- Argue that the model structure is wrong. Hitchcock (at PSA 2010) noted that antidote and poison have to mix in the body; if we represent that aggregating variable, the problem disappears. While solving the problem, we should like an independent account of when a model is adequate. Handfield et al. (2008) and Korb and Nicholson (2008) sketch an account using the idea of "possible processes."

We think that some combination of the last two responses—a mixture of an analysis of causal strength with a metaphysical analysis of causal processes—is the most promising framework for completing the story of actual causation.

5. Conclusion. Active path approaches treat actual causation in the same way one might treat type causation—assessing causal relevance by wiggling a potential cause and reading the changes in a downstream variable—except in applying a restricted context. The trick is to find the right context to do justice to our intuitions.

Hitchcock (2001) and Halpern and Pearl (2005a, 2005b) made some progress in this using deterministic criteria, that is, using only the Bayesian network structure and ignoring parameters. We have addressed that limitation, extending active path analyses to probabilistic models and defending the project against impossibility claims. Nevertheless, the resultant analysis is incomplete. For example, it cannot independently handle Hiddleston's superfluous antidote case. For such cases, we suggest considering an account that unifies difference making with causal process metaphysics, such as that presented in Handfield et al. (2008).

https://doi.org/10.1086/662957 Published online by Cambridge University Press

REFERENCES

Cheng, P. W. 1997. "From Covariation to Causation: A Causal Power Theory." *Psychological Review* 104 (2): 367–405.

Halpern, J. Y., and J. Pearl. 2005a. "Causes and Explanations: A Structural-Model Approach." Pt. 1, "Causes." *British Journal for the Philosophy of Science* 56:843–87.

———. 2005b. "Causes and Explanations: A Structural-Model Approach." Pt. 2, "Explanations." *British Journal for the Philosophy of Science* 56:889–911.

Handfield, T., C. R. Twardy, K. B. Korb, and G. Oppy. 2008. "The Metaphysics of Causal Models: Where's the Biff?" *Erkenntnis* 68:149–68.

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

Hitchcock, C. R. 2001. "The Intransitivity of Causation Revealed in Equations and Graphs." *Journal of Philosophy* 98 (6): 273–99.

Korb, K., L. Hope, A. Nicholson, and K. Axnick. 2004. "Varieties of Causal Intervention." In *Pacific Rim Conference on Artificial Intelligence*, ed. Chengqi Zhang and Hans Werner Güsgen, 322–31. New York: Springer.

Korb, K. B., and A. E. Nicholson. 2008. "The Causal Interpretation of Bayesian Networks." In *Innovations in Bayesian Networks: Theory and Applications*, ed. D. E. Holmes and L. C. Jain, 83–116. Berlin: Springer.

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

Twardy, C. R., and K. B. Korb. 2011. "Actual Causation by Probabilistic Active Paths (Supplement)." PhilSci Archive, http://philsci-archive.pitt.edu/8878/.

Woodward, J. 2005. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

https://doi.org/10.1086/662957 Published online by Cambridge University Press