Causation and contrast classes

Robert Northcott

Abstract I argue that causation is a *contrastive* relation: c-rather-than-C* causes e-rather-than-E*, where C* and E* are contrast classes associated respectively with actual events c and e. I explain why this is an improvement on the traditional binary view, and develop a detailed definition. It turns out that causation is only well defined in 'uniform' cases, where either all or none of the members of C* are related appropriately to members of E*.

1 Introduction

I shall argue that causation is a four-placed relation, so that any binary claim 'c causes e' should be glossed as:

c-rather-than-C* causes e-rather-than-E*

R. Northcott (✉)
Philosophy, University of Missouri-St Louis, 599 Lucas Hall (MC 73), One University Blvd, St Louis, MO 63121-4499, USA
e-mail: northcottr@umsl.edu

where C* and E* are *contrast classes*[1] associated respectively with actual events *c* and *e*.[2]

This approach is consistent with, indeed implied by, scientific practice with regard both to causal inference and to interventions. It is also consistent with, indeed implied by, a manipulationist view of causation such as that of Woodward (2003). I shall argue that we need to incorporate contrast classes when formulating any difference-making view of causation.

That choice of contrast class influences our causal talk is a familiar observation. The suggestion to be explored here, against most of the literature, will be that we should incorporate specification of contrast class into the semantics, rather than pragmatics, of causation. This will also enable us to present an account of the distinction between causation and causal *judgment*, and of the exact role in the latter played by context.

On this paper's view, causation is not a relation purely between actual events. Rather, we require evaluation not just of the relation between the actual events normally labeled 'cause' and 'effect', but also evaluation of counterfactual claims that are implicitly added when those events are described contrastively. Clearly, whether one event causes another will now thus be dependent on how those events are described. Nevertheless, I shall also argue that a contrastive approach can still do justice to our intuitions of causation's objectivity and mind-independence.

2 The two assassins puzzle

Hitchcock (2003) gives several examples where it seems that no current theory of causation can offer a convincing argument for whether an event is a cause or not, or for why we should even care. In order to motivate the contrastive view, consider how it helps here. I shall need to have a particular contrastive definition in mind. So, if O(*x*) denotes the occurrence of event *x*, then for actual events *c* and *e* and associated contrast classes C* and E*:

(D1) *c*-rather-than-C* causes *e*-rather-than-E* iff O(C*) counterfactually entails O(E*)

This is just a preliminary counterfactual definition for temporary use. Later, I develop a more refined version.[3]

I shall focus on the case (due originally to Michael McDermott) that Hitchcock himself highlights as the most intractable, concerning two assassins. Begin by presenting two different versions of the puzzle. First version: a Captain is training an Assistant in assassination. Only the latter has a gun. They are stalking a Victim in a crowded market

---

[1] Contrast classes are to be understood as non-empty sets of contrast events. For ease of exposition, I shall use 'contrast class' to refer to both singleton and non-singleton sets. I shall denote actual events by *c* and *e* and associated contrast classes by C* and E*. Later, I shall denote generically by C_0 and E_0 specific contrast classes that are members of the broader sets C* and E*. Where helpful, numerical suffixes such as C_1, E_2 etc will be used to denote different choices of C_0 and E_0.

[2] A contrastive approach is also ably defended by Schaffer (2005) and Maslen (2004), who apply it to the analysis of familiar paradoxes including transitivity and (in Schaffer's case) several others too. Two other important precursors are Woodward (1984) and Hitchcock (1996).

[3] For ease of exposition, often in what follows I shall follow normal usage by speaking of causation *as if* it held solely between two actual events, but in all such cases a specification also of contrast classes should be taken to be implied as appropriate. I shall argue in Sect. 6 that such linguistic ellipsis is in any case readily explicable and normally harmless. Also for ease of exposition, I shall frequently speak of counterfactual entailment holding between the contrast events themselves rather than between the propositions of their occurrence. Finally, except where stated otherwise, I examine only cases of token deterministic causation.

place when there occurs a great surge of people that threatens to carry Victim off to safety. Captain and Assistant become separated and lose visual contact. Assistant will not shoot without authorization from Captain. Therefore, in order still to have any chance of killing Victim before she gets away, Captain as an emergency measure yells to Assistant to 'fire!'. As a result, Assistant indeed fires. However, the noise of the yell is also heard by Victim who consequently ducks and as a result survives the shot before indeed escaping with the crowd. The question is: was Captain's yell the cause of Victim's survival? Intuitively, the answer seems to be 'no' since even if Captain had not yelled still Victim would have got away in any case. Thus Victim would likely be thanking the fortuitous surge of the crowd for enabling her to escape rather than thanking Captain for doing his best, under the circumstances, to *prevent* that escape.

Second version: suppose instead that there was no surge in the crowd and that in fact Captain and Assistant were standing at leisure on a balcony overlooking Victim, with plenty of time to select the moment to fire. Assume there is incentive not to let Victim get away without taking at least one shot at her. Captain could communicate to the eager Assistant at any moment by the prearranged signal of raising a finger. However, just as he is indeed about to raise his finger, Captain impulsively yells out loud 'fire!', which alerts Victim who consequently ducks and as a result survives the shot and escapes. Now again the question is: was Captain's yell the cause of Victim's survival? This time the answer seems to be 'yes', since this time the yell made all the difference. If only he had signaled silently as arranged, rather than bursting out into a yell, Captain could have ensured a successful assassination. That is, if only Captain had not yelled, the mission would not have failed and Victim would not have survived.

Any theory of causation framed—mistakenly in my view—purely in terms of actual events will inevitably carry a latent chink in its armor, and with sufficient cunning that chink can be deliberately exploited so as to engineer paradoxes. One trick is by careful presentation and unusual circumstance to leave crucially unclear just which contrast classes are implied. To see this, compare our two versions of the two assassins case to how the example is usually presented (Hitchcock 2003, 10):

> Two assassins, Captain and Assistant, are on a mission to kill Victim. Upon spotting Victim, Captain yells 'fire!', and Assistant fires. Overhearing the order, Victim ducks and survives unscathed ... let us now ask: did Captain's yelling 'fire!' cause Victim to survive?

Stripped of any detailed background story, the answer has now become intuitively unclear. On the one hand, the yell alerted Victim and so indeed enabled her to survive since if left unalerted she would surely (it is stipulated) have died. On the other hand, if Captain had not yelled then Assistant would never have fired in the first place and so Victim never been endangered, so the yell can hardly be held to have *caused* Victim's survival. Hitchcock points out the underlying threat-protector structure of the example: Captain's yell both initiates the threat to Victim (i.e. Assistant's shot) and also the mechanism protecting her (her overhearing and consequently ducking). He judges (11) that, in contrast to other threat-protector examples:

> The case of the two assassins is slightly different ... because the threat and the protector are equally indirect. It is this feature of the example that makes it so hard for our intuition to settle on a definite answer to the central causal question.

I sympathize with the spirit of this analysis and think that it can be expressed nicely using the contrastive account. Let c = Captain's yell, and e = Victim's survival. Specify a

contrast class for the effect term of E_1 = Victim died, and for the cause specify *two* possible contrast classes:

   C_1 = Captain gave no signal at all, i.e., continued quietly as before, and
   C_2 = Captain did give a signal, but a silent one

Given the specification of the problem, clearly C_2 counterfactually entails E_1, whereas C_1 does *not*. The answer to 'did *c* cause *e*?' is therefore that it depends critically on which contrast class is being invoked. With respect to *e*-rather-than-E_1, *c*-rather-than-C_2 is a cause whereas *c*-rather-than-C_1 is not.

   This divergence explains our split intuition in the case, since the story is (not by chance) framed in just such a way as to leave C_1 and C_2 the two main—and equally—plausible scenarios. Thus we are not sure which applies. Hitchcock reports that people's intuitions vary depending on how the example is presented to them. This may well be because subtle contextual cues are vital to indicating which of C_1 or C_2 is implicitly brought to mind. The best evidence for this in turn is actually to construct retellings of the example that direct our attention towards one or the other contrast class—which was of course the motivation for presenting the two different versions of the story earlier, and for noting the opposite intuitive reactions they generated.

   Hitchcock comments that since there is no dispute about what he calls the 'stage-one' substantive facts of the matter, subsequent disputes as to whether causation is 'really' present need not concern us. After all, if there are no other substantive disagreements then what difference does it make whether we declare Captain's yell a cause or not? I certainly agree that there is no causal fact of the matter with respect to the actual events alone. But I also believe we can see now that, nevertheless, whether we declare the yell to be a cause can matter after all, for it can serve to signal which contrast class is salient. Given context, that in turn may tell us much. For instance, here citing the yell as indeed the cause picks out C_2, i.e., our second scenario of Captain's incompetence. By contrast, claiming that the yell was *not* the cause of Victim's survival picks out instead C_1, i.e. the first scenario in which Captain behaved rationally and efficiently. Hence ascribing causation in this case suggests concisely that Captain was a blunderer and yelled unnecessarily, whereas not ascribing it suggests the opposite. I would bet that, when assessing responsibility, any court-martial of Captain would reason similarly.

   Further, altered circumstances could mean an ascription of causation having rather different implications. For instance, suppose it is now known in addition that Captain is an experienced officer fully aware that a yell would alert Victim. Then an ascription of causation might no longer license an inference of incompetence but rather now the different inference that by yelling unnecessarily Captain had for some reason been deliberately *warning* Victim. This then in turn might be thought to reveal either virtuous motives (at least from the point of view of Victim) or else perhaps some unknown self-interest, again depending on background knowledge regarding Captain's character, attitude towards Victim and so on.[4]

---

[4] Thus I disagree with Hitchcock's comments (2003, 10) that:

   ''We do not need an answer to this causal question to answer the standard barrage of questions. Does Captain deserve praise for saving Victim's life? Clearly he does not. If Captain actually wanted Victim to survive, did he pursue a rational course of action? No more rational than the alternative of withholding the order''.

But my view, as the main text argues, is that in some circumstances invoking causation might be very informative precisely with respect to questions such as these.

3 Relation to other theories of causation

The seeming relevance of contrast class to the truth values of causal claims has often been noted, with respect both to the cause slot (Hitchcock 1996, from which the Walter example below is adapted; Dretske 1977; Achinstein 1983) and effect slot (Van Fraassen 1980; Woodward 1984). For example, consider:

(1)   Walter's running to the classroom caused him to sweat

This causal claim seems incompletely specified. For consider each of:

(2)   Walter's running to the classroom, *rather than walking* to it, caused him to sweat
(3)   Walter's running to the classroom, *rather than cartwheeling* to it, caused him to sweat
(4)   Walter's running to the classroom, *rather than running to the bathroom*, caused him to sweat

Different specifications of contrast class apparently affect the truth values of causal claims: whereas (2) is plausible, (3) and (4) are less so.

A common rejoinder is that (1) merely answers the question of which event (i.e., Walter's running) caused the sweating, and does so perfectly determinately and correctly. On this view, (2), (3) and (4) are answers to a *different* question, namely which *aspect* of Walter's running caused the sweating? (Or, on some formulations, what *explains* Walter's sweating?) Perhaps, therefore, only the first of these two questions directs our attention to causation? If so, it could still be held that the causal relation is indeed a binary one between actual events. The 'rather than' formulation would then be seen merely as highlighting the causally efficacious or explanatory aspects of Walter's running rather than as picking out causation itself. But our analysis of the two assassins puzzle now indicates that, on the contrary, specification of contrast class is crucial when answering questions of the first sort too. In particular, choice of contrast class did not just determine which aspect of Captain's order caused Victim's survival; it also determined whether the event of Captain's order was the cause at all.

More generally, theories of causation commonly take *difference-making* to be essential—a cause is something that makes a difference with respect to its effect.[5] What are the advantages to capturing this via the formal device of contrast classes rather than in some other way? Begin a survey of the alternatives by considering counterfactual accounts, in particular those that apply a Lewis-style semantics. A general worry for all of these is how to interpret a counterfactual's antecedent. For example, if c = two people talking, then $\sim c$ might variously be thought to be them not talking, them talking differently, them talking at a different time or in a different place, different people talking, and so on. How do we know which? Merely asserting the negation of c does not determine causal properties fully. It might be thought that appeal could be made instead to some somehow 'natural' or 'normal' state, or to some general account of which features of an event are essential and which not. But unfortunately it does not seem that this can be achieved

---

[5] Exceptions include realist theories such as those of Tooley and Armstrong, and physical connection theories such as those of Salmon and Dowe. Although formally consideration of contrast classes could be grafted onto these too, there is more natural motivation to do so with difference-making accounts.

context-independently.[6] In the two assassins case, for instance, for c = Captain's yell, we saw: that '$\sim c$' could be either a silent signal or no signal at all; that context determined which of these options was salient; and in turn that that the choice was critical to determining whether c was a cause. In other words, an explicit specification of contrast class was crucial for a counterfactual account to give the right answer.

Of course, it might be argued in return that Lewis's apparatus of possible worlds automatically incorporates choice of contrast class already, since in any given context the appropriate interpretation of the negation is something that influences the selection of the nearest possible world—for that context. Thus: ''I'd suggest that ... contrasts ... are best understood as constraints on the range of relevant alterations'' (Lewis 2004, 99). ('Relevant alterations', in the context of Lewis's later influence account of causation, bears on precisely this issue.) But then it seems that the criteria for deciding similarity between possible worlds implicitly appeal to specification of contrast class. In particular, turning things around, the real driver of those similarity criteria starts to seem in part that they nominate the same nearest possible world that a contrastive definition would. That is, in order to work properly the criteria need to be granted a *precisification* of $\sim c$. So it is a necessary theoretical improvement to incorporate that precisification explicitly.

Similar remarks apply to counterfactual versions of probabilistic causation, whereby causation is defined as token chance-raising relative to the chance in a counterfactual world. Again, the worry concerns whether we can interpret the relevant counterfactuals as desired without explicit appeal to a contrast class.

When we move on to look instead at difference-making theories *not* based on Lewisian counterfactuals, we typically find that an explicitly contrastive structure has *already* been adopted—thus endorsing the approach of this paper. Begin with counterfactual accounts again. A prime example of a non-Lewisian approach here is Woodward's (2003) manipulationist theory of causation, which although concerned primarily with type-level claims, also implies a contrastive view of token causation. In his account, causation is defined in terms of idealized *interventions* on a causal structure. (These interventions are 'idealized' in that they must be assumed, roughly speaking, not to impact on an effect variable via any route other than through the cause variable, and also must be assumed to leave the relations comprising the causal structure—as opposed to the particular variable values inputted into it—unchanged.) Any such intervention represents a change in the value of a cause variable, and the impact of that on an effect variable can then be calculated from the causal structure. Denoting the post-intervention value of the cause variable as c and its pre-intervention value as C_0, and the post- and pre-intervention values of the effect variable likewise as respectively e and E_0, the resultant definition of causation then appeals (from the standpoint of after the intervention) to exactly the same counterfactual C_0–E_0 entailment as does our own definition (D2) below. Thus no intervention can be specified without appealing, in our terminology, to contrast classes. ''All causal claims must be interpretable as having a contrastive structure'' (Woodward 2003, 146; see also 67–8, 210–15).[7]

---

[6] Humphreys (1989, 38) defines the *neutral level* of a cause to be that level at which ''the property corresponding to that variable is completely absent''. But although useful for other purposes, this cannot help us here. For such a neutral level in effect already presupposes some specification of $E*$ ('the property ... completely absent'). In our terminology, the specification of $C*$ is objective only once given an $E*$, so there inevitably still remains one degree of freedom, as it were, for contextual considerations.

[7] As Woodward remarks, on such a view the causal relata can be viewed as *changes* in the values of variables. More precisely, the relevant changes here are those between c and C_0, and between e and E_0.

Woodward's manipulationist approach borrows extensively from the recent literature in causal modeling (Spirtes et al. 2000; Pearl 2000). The evaluation of counterfactuals there, as in Woodward, is via appeal to causal models rather than to Lewis's nomologically-based similarity criteria. Clearly, this approach therefore does not offer a reductive analysis of causation. But in return it is arguably much better able to handle the interpretive difficulties with a Lewisian approach mentioned earlier, in effect by incorporating contrast classes explicitly. It is also better able to handle counterfactuals with complex antecedents, i.e., to evaluate what happens after (in manipulationist terminology) multiple interventions that alter more than one variable. Such cases, which are analogous to the 'group' causation mentioned below in footnote 16, again currently seem impossible to tackle without, among other things, an explicit specification of contrast classes (Woodward 2003, 136–145).

Turn next to non-Lewisian theories of probabilistic causation. These are typically formulated in terms of conditional probabilities of type events rather than counterfactual chances of token ones. (Hitchcock 1996) argues persuasively that these probabilities too are best formulated contrastively. For example, smoking 20 cigarettes per day increases the probability of lung cancer relative to smoking 10 per day, but decreases it relative to 40 per day. Formally, event type C is a (contributing) cause of event type E (once, roughly speaking, all other causal factors have been controlled for) iff: $pr(E \mid C) > pr(E \mid C^*)$. Analogously to before, the causal properties of $\sim C$ need to be disambiguated by invocation of a specific contrast class C^*.[8] Indeed, to this end the two assassins example could be re-told, now as a type-probabilistic story—suppose Captains' yells merely increase the probability of Assistants firing rather than making it certain. Then whether such yells are causes of Victims surviving will again depend on choice of contrast class, for exactly the same reasons as before.

As Woodward highlights at length, there is also an intimate connection between the contrastive conception of causation and scientific practice. For example, the method of controlled experiment—the gold standard of causal inference if anything is—is designed precisely to assess the impact on an effect (i.e., e starting from some E_0) of changing a causal input (i.e., of c relative to C_0). Note that this impact will of course therefore vary depending on the choice of C_0 and E_0. Moreover, the whole point of controlling for other causes is to ensure that, as far as possible, our actual data are a good proxy for the data that *would* have been generated in the ideal (counterfactual) situation of exactly identical background conditions, that is to say in the 'nearest possible world' in which C_0 and E_0 reside. In other words, the method of controlled experiment is designed to enable us as best we can to evaluate from actual data exactly the counterfactual that the contrastive view promotes as salient (Northcott 2006). (Thus, notwithstanding its incorporation of 'subjective' pragmatic factors, far from a contrastive approach being antithetical to objective scientific inquiry, it is in fact endorsed by it.)

Finally, attributions of causation in ethics and in the law are also sensitive to contrastive considerations (Hart and Honore 1985). Often so too are attributions in everyday life. And while everyday attributions are admittedly also often binary in form, a contrastive account offers a good explanation of that latter phenomenon too (Sect. 6).

---

[8] I omit discussion here of how we might incorporate a contrastive sensibility on the effect side too.

4 A definition

Recall our preliminary definition from earlier:

> (D1) c-rather-than-C^* causes e-rather-than-E^* iff $O(C^*)$ counterfactually entails $O(E^*)$

In keeping with the contrastive emphasis on difference-making, (D1) is a counterfactual definition. In this paper, I shall be fleshing out nuances specific to a contrastive version of such a definition, not addressing general objections (such as overdetermination cases) applicable to all counterfactual definitions alike.

As it stands, the above definition seems unconstrained by c and e on its right-hand side. I return to this below. Also, as Kim pointed out long ago (1973), for c to be a cause of e it is necessary that c and e be distinct events else writing 'tri' could be deemed a cause of writing 'trivial'.

Next, turn to the issue of constraints on the specification of contrast class. By this, I mean the constraints that contrast classes must obey *if* causation is to be ascribed. First, it is necessary that C^* and E^* be able to stand in the counterfactual entailment relation, which will presumably hold (given the discussion below) for virtually any c occurring in e's back lightcone.

Second, since c and e are events in the actual world it is necessary (for c to cause e) that C^* and E^* do *not* also occur in the actual world, i.e., that they be counterfactual. Thus if $e = E^*$, for instance, it follows automatically that c cannot be a cause of e.[9] But while necessary, this condition is not sufficient since it is so weak as to allow through virtually any c-e pair. For instance, for $C_1 = $ Walter walks to the classroom and $E_1 = $ he maintains normal skin moisture, then $c = $ I had dinner last night would come out as a cause of $e = $ the sun rose today. Intuitively it seems that contrast classes must, as it were, indeed genuinely contrast. But requiring that C^* *contradict* c (and E^* contradict e) seems, on the other hand, too strong. For example, if there is some possible world in which Walter can both walk and run (via time travel, maybe) then the criterion would rule out the C^* contrast class that we used in this very example.[10] So instead I propose here: that c and C^* (and similarly e and E^*) must be physically incompatible, i.e., that they co-exist in no possible world with the same laws as ours.

For any given c and e, this nomological criterion determines respective *sets* $\{C^*\}$ and $\{E^*\}$ of eligible contrast events. Up to now, I have glossed over the need to express a contrastive account of causation in terms of particular contrast events rather than these broad contrast sets $\{C^*\}$ and $\{E^*\}$. But for convenience, from here on label by C_0 and E_0 specific members of $\{C^*\}$ and $\{E^*\}$ respectively.[11] Then we may write a revised version of our definition:

---

[9] Formally, proscribing $e = E^*$ is equivalent to Woodward and Hitchcock's demand that causation be defined only via '*testing* interventions' (Woodward and Hitchcock 2003, 17). Informally, it is just the demand that any cause make a difference.

[10] Some Lewisian formulation such as requiring, roughly speaking, that $C^*\&E^*$ obtain in a closer possible world than any in which either of $c\&C^*$ or $e\&E^*$ do, will be too weak again. For this does not imply counterfactual entailment between C^* and E^*—for instance, because the conjunction of C^* and some *other* contrast class, E_1 say, might be in a yet closer possible world even than $C^*\&E^*$. I am also reluctant to build in from the start a commitment to any particular semantics for counterfactuals, Lewisian or otherwise.

[11] What determines the selection of these C_0 and E_0? In brief, the answer is *context*. It will follow that our causal judgment is correspondingly context-sensitive, just as in the two assassins case (Sect. 6).

(D2) c-rather-than-C_0 causes e-rather-than-E_0 iff $O(C_0)$ counterfactually entails $O(E_0)$[12]

The actual events c and e do still constrain the right-hand side of (D2). To see this, notice that the sets $\{C^*\}$ and $\{E^*\}$ themselves yield a set of possible (C_i, E_j) pairs, where C_i and E_j denote individual contrast events as stated earlier. For some such pairs, C_i will counterfactually entail E_j, for others it will not. Thus each (C_i, E_j) pair is in turn a member of one of two further sets: either $\{G\}$ = the set of such pairs where C_i does counterfactually entail E_j; or $\{B\}$ = the set of pairs where it does not. $\{G\}$ and $\{B\}$ ('Good' and 'Bad') form an exhaustive disjoint partition of the set of all (C_i, E_j) pairs. So we may re-phrase (D2) as follows:

(D3) c-rather-than-C_0 causes e-rather-than-E_0 iff (C_0, E_0) is a member of $\{G\}$

Given the nomological constraint on selection of C^* and E^*, $\{G\}$ and $\{B\}$ are themselves constrained by c and e (as well as by the prevailing conditions).

5 Non-singleton contrast classes—when is causation well defined?

What happens when our particular choices of contrast class—i.e., C_0 and E_0—are themselves non-singleton? Consider an example. Let c = a politically strong defense minister takes office, e = the military budget is set at 1 bn dollars, $\{C_0\}$ = the *set* of salient alternative defense ministers C_i each taking office, and $\{E_0\}$ = the set of military budgets of less than 1 bn dollars. Suppose the claim is that—the strong minister taking office was the cause of a higher budget than otherwise. There are three possibilities. First, if all the C_i counterfactually entail corresponding $E_i < 1$ bn, then there is no problem: clearly, relative to the alternative ministers $\{C_0\}$, the actual minister taking office was a cause of a higher budget since all of the alternatives would have obtained only a lower budget. Second, likewise if all the C_i counterfactually entail $E_i > 1$ bn, then the answer to the causal claim is unambiguously negative.

The situation becomes murkier only in the third and general case where some of the C_i counterfactually entail $E_i < 1$ bn and others $E_i > 1$ bn. Was the defense minister a cause of a higher budget, given that some of the alternatives would have obtained a lower budget but others a higher one? Even intuitively, without further information the given causal claim now seems impossible to evaluate. (Readers are invited to try for themselves.) The situation does improve, though, if one of the counterfactual outcomes achieves superior saliency. For example, suppose we are told that the alternative minister was 'probably' one who would have obtained a budget of under 1 bn. Then it seems we are now correspondingly inclined to declare the actual minister taking office 'probably' the cause of a higher budget. Notice though that by assumption the world in this case is deterministic; thus the 'probably' here must be epistemic. In particular, the new information therefore merely reduces our uncertainty concerning which determinate fact of the matter is salient; it does not alter that causation is only well defined when there *is* some such determinate fact of the matter.

The two assassins puzzle, when presented so as to leave C_0 ambiguous, is another example of this third type of case. The ambiguity there ensured that the set of salient

---

[12] If C_0 and E_0 are taken to be respectively c and e's negations (more strictly, the negations of the propositions of their existence), formally (D2) reduces to Lewis's original definition of causal dependence (1973). But recall our earlier caveat that the causal properties of such negations are in general underdetermined save by smuggling in contrastive considerations.

contrasts included both a C_i that did, and a C_i that did not, imply causation, and further that there was no decisive reason to deem either more salient than the other. As a result, it was impossible to evaluate whether c was a cause. But as we saw in Sect. 2, when one of the C_i does achieve dominant saliency, so causation becomes clear.

I conclude that causation is only even well defined with respect to what we might informally label 'uniform' $\{C_0\}$ sets, i.e., sets where either all or none of $\{C_0\}$'s component C_i counterfactually entail an E_i in $\{E_0\}$. (Thus a singleton C_0 meets this uniformity criterion automatically.[13]) Indeed, the fundamental need for explicit specification of contrast classes only arises in the first place because of exactly this fact. In particular, the need to disambiguate the negations '$\sim c$' or '$\sim e$' arises precisely when, as in the two assassins puzzle, they denote non-uniform sets of contrasts.

The upshot is that:

1. When C_0 is non-singleton, c causes e iff all the C_i in the specified $\{C_0\}$ counterfactually entail E_0.
2. When E_0 is non-singleton, c causes e iff C_0 counterfactually entails any E_i within the specified range of $\{E_0\}$.[14]

Summarizing, we may write a generalized definition (for the deterministic case). For two distinct actual events c and e with respective associated contrast classes $\{C_0\}$ and $\{E_0\}$:

(D4) c-rather-than-$\{C_0\}$ causes e-rather-than-$\{E_0\}$ iff: for all C_i in $\{C_0\}$, $O(C_i)$ counterfactually entails $O(E_i)$ such that E_i is in $\{E_0\}$.[15]

Clearly, for the case when $\{C_0\}$ and $\{E_0\}$ are singleton this just reduces to (D2). The constraints on choice of contrast class will also be analogous to before: in particular, necessary conditions for c to cause e will be that c is not a member of $\{C_0\}$ and e not a member of $\{E_0\}$.[16]

---

[13] It does so, at least, as long as we assume determinism. Otherwise, a single C_0 might lead to either E_1 or E_2, where only one of the latter is in $\{E_0\}$. For example, suppose that an alternative defense minister taking office could chancily have led to a budget of either 0.5 or 1.5 bn. For the first figure, the actual minister taking office was therefore the cause of a higher budget, for the second the cause of a lower one. In this now indeterministic world, causation would then presumably depend on the two outcomes' relative *probabilities*. But this paper will concentrate only on the deterministic case.

[14] Notice that *continuous* contrast classes, such as $\{E_0\}$ in the defense minister example, do not seem to pose any particular difficulty. The conditions for causation can be taken to apply simply to every element in the relevant set, regardless of whether those elements are countable.

[15] Schaffer (2005, 318) proposes an extra requirement, namely that every E_i in $\{E_0\}$ must be entailed by some C_i in $\{C_0\}$. But I think this addition is unnecessary. Consider the case of the defense minister in the text, for example. Clearly, the minister taking office is still a cause of a higher budget even if there exists some budget figure below 1 bn dollars that, it turns out, would have been obtained by none of the possible alternative ministers. There is an asymmetry here between cause and effect: whereas every C_i in $\{C_0\}$ must entail an appropriate E_i, not every E_i in $\{E_0\}$ must be entailed by an appropriate C_i. This asymmetry, it seems to me, is just reflecting the logical asymmetry of the entailment relation: 'X entails Y' requires that all X's must indeed entail Y's, but not that all Y's must be entailed by some X.

[16] Perhaps some account of 'group' causation could also be attempted by formulating a definition for when c and e are themselves non-singleton. Such a definition would be necessary for assessing cases of, in manipulationist terminology, complex interventions (Sect. 3). Let $\{c\} = \{c_1, \dots, c_n\}$ and $\{e\} = \{e_1, \dots, e_m\}$, and let the respective contrast classes, one for each element of $\{c\}$ and $\{e\}$, be denoted by $\{C_0\} = \{C_1, \dots, C_n\}$ and $\{E_0\} = \{E_1, \dots, E_m\}$. Then my initial suggestion would be: $\{c\}$-rather-than-$\{C_0\}$ causes $\{e\}$-rather-than-$\{E_0\}$ iff $O(C_1 \, \& \, \dots \, \& \, C_n)$ counterfactually entails $O(E_1 \, \& \, \dots \, \& \, E_m)$. This new definition could itself be generalized, in the manner of (D4), to cases where individual c_i and e_i have non-singleton contrast sets.

6 Context and causal judgment

We have reached the following picture: metaphysically, causation is defined on 4-tons (c, e, C_0, E_0). For any given such 4-ton, whether c causes e depends on whether C_0 and E_0 are members of $\{C^*\}$ and $\{E^*\}$ respectively, and also on whether C_0 counterfactually entails E_0. Assuming the objectivity of nomological incompatibility and of counterfactual entailment, causation too is thus rendered perfectly *objective*.

One problem with our definition is that it endorses many cases of causation that seem intuitively dubious. For example, consider c = the book is in the open air, e = the book is being read by me, C_0 = the book is buried in the ground, E_0 = the book is hidden from the light. Then the 4-ton (c, e, C_0, E_0) is endorsed as causal, since as required C_0 counterfactually entails E_0, and c and e are nomologically incompatible with C_0 and E_0 respectively. The answer here, I think, is that our causal *judgments* are sensitive only to those contrast classes relevant in the *context*. That is, even though the definition may endorse many different C_0–E_0 pairs for any given c-e pair, context selects out for our causal judgment's attention only a subset of those C_0–E_0 pairs.[17] Thus the sometime mismatch between the verdicts of the definition and those of our causal judgment. What is necessary now is to give some account of just how it is that certain C_0–E_0 pairs are prioritized for our attention in this way. In addressing this, I shall first take a detour into our use of causal language.

An obvious starting point is that when natural language speaks of causal relations holding between actual events c and e, it often makes no explicit reference to contrast classes at all. It is thereby left underspecified exactly which C_0–E_0 pair is being invoked and thus which counterfactual entailment is being asserted. Nevertheless, in practice this inexplicitness of natural language *rarely matters*, principally because it is usually obvious from the circumstances which C_0–E_0 pair (or set of such pairs) is being specified implicitly. For example, for c = my chop with the axe, and e = the wooden log breaks, it may well be obvious that the salient contrast classes are:

C_1 = I left the axe on the ground (such a contrast event is itself often only specified obliquely via the invocation of a negation like 'I did not chop'), and
E_1 = the log stayed intact.

In experimental contexts also, usually it is obvious what the baseline contrast level is. For example, we might test the impact of strengthening an electromagnetic field—in which case, the previous field strength and level of effect variable are the obvious contrasts. The choice of C_0 and E_0 is understood by all from the start.

Maudlin (2004) offers general reason to think that a context will tend to suggest a specific contrast automatically. On his account, special science generalizations typically stipulate a default state of affairs, any deviation from which requires explanation. In this respect, he sees them as emulating Newtonian mechanics, whose first law stipulates the default motion of inertia, any deviation from which must then be explained by a force. Causation is then seen as our label for tracking and explaining such deviations. If by 'context' we include the generalization by which we are analyzing a case, then it is clear how context determines contrast classes—they are just what would have occurred in the default, or 'inertial', case. Hitchcock (2007) emphasizes a similar default/deviation

---

[17] Note that the nomological stipulation that generates $\{C^*\}$ and $\{E^*\}$ is independent of the operation of context. In principle, context is thus also free to select C_0 and E_0 that are not members of respectively $\{C^*\}$ and $\{E^*\}$, in which case we would not have causation.

dichotomy. In his scheme, causal models always carry with them a specification of a default state to which our judgment is then sensitive when interpreting them.

Returning to the question of just how context determines choice of contrast class, we thus see that the very formulation of causal claims may automatically carry built-in, so to speak, a default contrast scenario. On the other hand, it is too strong to claim that there exists an objective procedure for nailing down contrast class in every case.[18] As Maudlin himself points out, and Hitchcock concedes too, often, at least in special sciences, we have a *choice* of generalizations or models with which to analyze a case. This may impact in turn on choice of contrast class and thus, sometimes, make all the difference to causation. Take the two assassins case, for instance. As it is usually presented, more than one generalization is potentially salient. Thus, in the first version from Sect. 2, the salient generalization might be that visual signaling through a crowd is impossible, generating a default scenario of Captain giving no signal. But in the second version, set on a leisured balcony, now the salient generalization might be that competent Captains signal silently, generating that as the default scenario. As we saw, opposite consequences follow for whether Captain's yell is a cause. In Hitchcock's terms, the usual presentation of the story leaves it ambiguous which causal model is appropriate—one in which the default value of the variable representing Captain's behavior is a silent signal or no signal at all.

Even when circumstances or presentation do not leave it sufficiently obvious which contrast classes are appropriate, a wide array of linguistic devices are available precisely to remedy this problem and to ensure that it is indeed always clear which C_0–E_0 pair a speaker has in mind. One familiar such device is *linguistic stress*. Suitable use of this can easily disambiguate which contrast classes are appropriate. For example, returning to our earlier example, announcing that the cause was 'I chopped with the *axe*' while a hammer is also lying handy might indicate:

C_2 = I chopped with the hammer,

while emphasizing that '*I* chopped with the axe' could imply instead:

C_3 = my friend did the chopping.

Another device is *choice of vocabulary*: some words connote different contrasts than do others. No doubt there exists myriad other subtle devices too, the exact nuances of whose use is highly context-specific.

Overall, as a result of this rich store of linguistic practice, it follows that even though on the contrastive view '*c* causes *e*' necessarily makes implicit reference also to contrast events C_0 and E_0, still we can usually talk without confusion *as if* causation were indeed solely a relation between the actual events *c* and *e*.[19] If really necessary, finally, clarification can in any case always be made explicit, for instance by using the 'rather than' locution or similar: 'I chopped with the axe *rather than* wait for woodworm to strike.' In conclusion, therefore, inexplicit specification of contrast classes is typically a problem that is easily surmounted. So this lack of explicitness in natural language is not costly in everyday life—which, given also its convenience, perhaps explains its survival.

---

[18] As well as Maudlin and Hitchcock, see also Schaffer (2005), Maslen (2004) and Van Fraassen (1980) for further discussion.

[19] Indeed, so easily can this be done that it is perhaps thereby explained why it could become natural even to *believe* that causation is a relation solely between actual events—such mistaken reification is hardly unprecedented in human psychology. In principle, the contrastive view thus has a ready explanation for this mistaken belief, and accordingly that belief's existence loses force as a counterargument.

Finally, if no fully objective procedure for how context determines contrasts is available, then how can we be sure that any particular invocation of contrast classes is well motivated? What constrains appeal to context? One test is to manipulate the context *first*, and then see whether our causal judgments indeed do follow along in the way predicted (Van Fraassen 1980). In lieu of a fuller empirical investigation, the success of such a manipulation is then the best evidence for the claim that context is indeed doing the independent work claimed for it. And the analysis of the two assassins case in Sect. 2 was precisely an example of this: upon manipulating context, our causal judgments indeed changed as predicted.

References

Achinstein, P. (1983). *The nature of explanation*. Oxford: Oxford University Press.
Collins, J., Hall, N., & Paul, L. A. (Eds.) (2004). Causation and counterfactuals. (Massachusetts: MIT Press.).
Dretske, F. (1977). Referring to events. In P. French, T. Uehling Jr & H. Wettstein (Eds.), *Midwest studies in philosophy II* (pp. 90–99). Minneapolis: University of Minnesota Press.
Hart, H. L. A., & Honore, A. (1985). *Causation in the law*. Oxford: Clarendon Press.
Hitchcock, C. (1996). The role of contrast in causal and explanatory claims. *Synthese*, *107*, 395–419.
Hitchcock, C. (2003). Of humean bondage. *British Journal for the Philosophy of Science*, *54*, 1–25.
Hitchcock, C. (forth coming). Prevention, preemption, and the principle of sufficient reason. Philosophical review.
Humphreys, P. (1989). *The chances of explanation*. Princeton, NJ: Princeton University Press.
Kim, J. (1973). Causes and counterfactuals. *Journal of Philosophy*, *70*, 570–572.
Lewis, D. (1973). Causation. *Journal of Philosophy*, *70*, 556–567.
Lewis, D. (2004). *Causation as influence*. In J. Collins, N. Hall, & L. A. Paul (Eds.), (pp. 75–106), Cambridge, Mass: MIT Press.
Maslen, C. (2004). Causes, contrasts and the nontransitivity of causation. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 341–357). Cambridge: MIT Press.
Maudlin, T. (2004). Causation, counterfactuals, and the third Factor. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 419–443). Cambridge, MA: MIT Press.
Northcott, R. (2006). Casual efficacy and the analysis of variance. *Biology and Philosophy*, 21(2), 253–276.
Pearl, J. (2000). *Causality*. New York: Cambridge University Press.
Schaffer, J. (2005). Contrastive causation. *Philosophical Review*, *114.3*, 297–328.
Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search (2nd edition)*. Cambridge, Massachusetts: MIT Press.
Van Fraassen, B. (1980). *The scientific image*. Oxford: Oxford University Press.
Woodward, J. (1984). A theory of singular causal explanation. *Erkenntnis*, *21*, 231–262.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.
Woodward, J., & Hitchcock, C. (2003). Explanatory generalizations, Part I: A counterfactual account. *Nous*, *37*, 1–24.