Philos Stud
DOI 10.1007/s11098-009-9392-8

Natural-born determinists: a new defense of causation as probability-raising

Robert Northcott

Abstract A definition of causation as probability-raising is threatened by two kinds of counterexample: first, when a cause lowers the probability of its effect; and second, when the probability of an effect is raised by a non-cause. In this paper, I present an account that deals successfully with problem cases of both these kinds. In doing so, I also explore some novel implications of incorporating into the metaphysical investigation considerations of causal psychology.

Keywords Causation · Probability · Psychology · Determinism · Ex post · Ex ante

1 Introduction

The central idea of probabilistic causation is that a cause is something that increases the probability of its effect. It is clear immediately that two kinds of counterexample are possible, threatening respectively this criterion's necessity and sufficiency: first, when a cause lowers the probability of its effect; and second, when the probability of an effect is raised by a non-cause. In this paper, I present an account that deals successfully with problem cases of both these kinds.

The account has a peculiar feature though, namely that according to it something is deemed a cause not because it raises the probability of its effect there and then, as is usually proposed, but rather only because it raises the probability of its effect *later* in time. However, although this formal condition is indeed peculiar metaphysically, I shall argue that it has a very natural interpretation psychologically—as a symptom of the fact that, at least with regard to causal judgment, we humans are committed *determinists*. This leaves us with an interpretive choice. Either, first, the formalism

R. Northcott (✉)
Department of 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

Published online: 24 March 2009

to be developed below can be read as a straightforwardly metaphysical account of probabilistic causation. Or else, second, it can be read instead as a theory of our causal judgment rather than of causation itself.

The formal analysis of the core problem cases stands on its own merits, and is compatible with either option. Nevertheless, I shall frame the discussion with the second interpretation in mind. In the final section, I explain why—the reason is that even on a psychologistic reading the account still turns out to have metaphysical implications; and, further, those implications are now of a more desirable kind. In particular, I shall argue that probability-raising at a later time, when understood as a theory of causal judgment, actually provides support for the more traditional metaphysical view of causation as probability-raising there and then. That in turn is desirable for reasons beyond mere lack of peculiarity. For example, it thereby saves from counterexample the view of causation already central to the contemporary causal modeling literature, causal decision theory, and scientific practice generally. The psychological detour thus pays metaphysical dividends. The exact argument for how it does so is methodologically novel. But to prepare the ground for all that, first it will be necessary to develop and apply the formal account in detail.

2 Ex ante versus ex post

Begin by considering token cases. On the standard view, probability is to be understood in them as objective single-case chance and the probability-raising criterion becomes a species of token counterfactual account—a cause is something that increases the chance of its effect compared to what that chance would have been had the cause been absent.[1] I shall assume, in common with most of the relevant literature, that the causal relata are events. Nothing important hangs on this though, as analogous eventual conclusions could still be drawn even given other choices of relata such as facts or aspects. In all our examples, *c* and *e* will denote actual events, *c* occurring before *e*, and we shall be concerned with the objective chance of *e* occurring.

Now turn to what will prove a key distinction—that between what I shall label *ex ante* and *ex post* probability. I shall take 'ex ante' to refer to the time when *c* occurs or, in those counterfactual cases where *c* does not occur, to the time in the nearest $\sim c$-world corresponding to when *c* occurs in the actual one (where '$\sim c$' denotes not-*c*). I shall take 'ex post', on the other hand, to refer to the time when *e* occurs or, in those counterfactual cases where *c* does not occur, to the time in the nearest $\sim c$-world corresponding to when *e* occurs in the actual one.[2] It follows that the

[1] For example Lewis (1986). I omit discussion of exactly how to understand single-case objective probabilities on the grounds that, with regard to the issues discussed in this paper, that is not the salient locus of philosophical dispute. I discuss the relevant counterfactuals at greater length in due course. Finally, for more on interpreting 'absent', see footnote 6 below.

[2] If *c* is a hastener or delayer, then the counterpart of *e* in the nearest $\sim c$-world will occur at a slightly different time to *e* in the actual world. In those cases, each ex post probability should be evaluated at the time *e* occurs in its own world. In particular, $p_c^{post}(e)$ and $p_{\sim c}^{post}(e)$ will therefore be evaluated at slightly different times. (I do not address here the vexed issue of how best to individuate events, and thus of just when a change in the timing of *e* is so great as to mean that the *e* in the nearest $\sim c$-world should be counted a different event altogether, and thus $p_{\sim c}^{post}(e)$ be declared zero.)

chance of *e* evaluated ex post is trivially always either 0 or 1, depending solely on whether *e* also occurs. (This follows immediately from, for instance, Lewis's definition of chance.) Thus the ex post chance of *e* is really a 'probability' at all only in a trivial formal sense, and might equally be thought of instead merely as a binary marker of whether *e* occurs. Moreover, it is trivially 0 or 1 both in the actual world where *c* does occur and also in the nearest world in which *c* does not, although how the latter chance is *perceived* by us is a more complicated matter, as we shall see.

Let $p_c(e)$ denote the chance of *e* occurring in the actual world, and $p_{\sim c}(e)$ denote the chance of *e* occurring in the nearest $\sim c$-world. Then the criterion for *c* being a probabilistic cause of *e* may be written: $\mathrm{p_c}(e) - \mathrm{p_{\sim c}}(e) > 0$. In words, *c* is a cause of *e* iff it makes *e* more likely, i.e. iff $\mathrm{p_c}(e)$ is bigger than $\mathrm{p_{\sim c}}(e)$.

In almost the entire literature these probabilities have been understood to be, in our terminology, ex ante. That is, attention has always focused on the chance of e, evaluated at the time *c* occurs. Here, I propose that we should instead frame the criterion in terms of probabilities understood to be *ex post*.[3] Label the ex ante probabilities $\mathrm{p_c^{ante}}(e)$ and $\mathrm{p_{\sim c}^{ante}}(e)$, and the ex post ones $\mathrm{p_c^{post}}(e)$ and $\mathrm{p_{\sim c}^{post}}(e)$. Then, for c and e distinct actual events, c occurring before e, the proposed new criterion is:

$$c \text{ cause } e \text{ iff } \mathrm{p_c^{post}}(e) - \mathrm{p_{\sim c}^{post}}(e) \qquad \text{[EP]}$$

That is, I propose that the relevant inequality should be with respect to ex post, rather than ex ante, probability.[4] Intuitively, the idea is that in probabilistic cases we judge only *in hindsight*, deeming something a cause depending on whether its chancy effect actually did subsequently occur (and also on whether its chancy effect *would* have occurred in nearby possible worlds).

---

[3] Note that this criterion is not a final *analysis* of causation. For one thing, as we shall see (Sect. 7), it seems that the relevant counterfactuals cannot be evaluated reductively. For another, this paper does not address another important class of problem cases, namely those of overdetermination (although Sect. 9 will mention one strategy for doing so).

[4] This proposal is entirely different from that advocated by Kvart (2004), notwithstanding the superficial resemblance between Kvart's notion of 'ex post facto' causation and mine of ex post probability. Kvart's ingenious idea is designed to yield the probability-raising result even for the ex ante probabilities, by stipulating that they be conditionalized on *possible* intermediate events between cause and effect. These intermediate events could raise or lower the ex ante probability. An ex post probability of our kind, on the other hand, in effect conditionalizes only on *actual* intermediate events (or, in counterfactual instances, only on those intermediate events occurring in the nearest $\sim c$-world). More particularly, it makes no *necessary* reference to intermediate events. As a result it can be applied even to cases without such events, and thus avoids counterexamples to Kvart associated with his scheme's need to invoke them (e.g. Dowe 2004, pp. 33–34). Moreover, our approach has no ambition to define (token) causation in terms of ex ante probability.

The proposed criterion is also distinct from Eells's (1991) interesting account of token causation in terms of probability trajectories. Among other differences, our criterion takes no account of how the chance of *e* evolves between *c* and *e*, but rather only considers its final value at the time of *e*.

The closest predecessor in the literature is a suggestion due to Ned Hall, as reported by Christopher Hitchcock (2004, p. 414). Hall suggests that we evaluate the probability of an effect 'shortly before the time at which the effect occurs.' Hitchcock also outlines a related proposal of his own, offered, like Hall's, as a solution to the Two Bullets case (Example Three below). But as Hitchcock goes on to explain, both these proposals suffer from having to assume certain probabilistic details about the particular Two Bullets case, plus also from having to assume that certain ex ante probabilities can take values of 0 and 1 (thus committing themselves a priori to denying that the universe is indeterministic 'all the way down'). This paper's account, by evaluating the probability at the time of *e* itself, avoids both these defects.

Given our definition of ex post probability, [EP] is equivalent to a simple counterfactual account—for actual events *c* and *e*, *c* is a cause of *e* iff O($\sim c$) entails O($\sim e$).[5] As will become apparent, I have framed the account in terms of ex post probability in order to elucidate its applicability to probabilistic cases, and also to emphasize what turns out to be a highly salient distinction—that between causal judgment and the metaphysics of causation itself.

The new criterion [EP] is easiest to appreciate by means of example. Start with a case that is relatively simple and uncontroversial.

Example One: Indeterministic Bomb. An unstable atom is placed in a box wired up such that if the atom decays then a bomb will be triggered. There is no other way for the bomb to be triggered, so it explodes if and only if the atom decays. Suppose that the device is due to be disconnected after one day, and that the atom has a 0.5 chance of decaying in this period. Suppose finally that, in fact, the atom does decay and so the bomb does indeed explode.

Let $c =$ the event of placing the atom in the apparatus, $\sim c$=not doing so,[6] and $e =$ the bomb's explosion. A comparison of the ex ante probabilities yields:

$$\mathrm{p}_{\mathrm{c}}^{\mathrm{ante}}(e) - \mathrm{p}_{\sim\mathrm{c}}^{\mathrm{ante}}[e] = 0.5 - 0 = 0.5 > 0.$$

Thus *c* increases *e*'s ex ante probability, so on the standard view *c* is endorsed as *e*'s cause, and all seems well.

What of our new approach here? First, since it was stipulated that the bomb did in fact explode, we know immediately that $\mathrm{p}_{\mathrm{c}}^{\mathrm{post}}(e) = 1$. What of $\mathrm{p}_{\sim\mathrm{c}}^{\mathrm{post}}(e)$? Here we must evaluate an 'ex post counterfactual.' The obvious evaluation in this case is: $\mathrm{p}_{\sim\mathrm{c}}^{\mathrm{post}}(e) = 0$. The reasoning is that, on the supposition that $\sim c$ transpired, by assumption there was no chance then that *e* also transpired. (More on ex post counterfactuals later.)

So criterion [EP] yields a calculation of:

$$\mathrm{p}_{\mathrm{c}}^{\mathrm{post}}(\mathrm{e}) - \mathrm{p}_{\sim\mathrm{c}}^{\mathrm{post}}(\mathrm{e}) = 1 - 0 = 1 > 0.$$

Thus [EP] too endorses *c* as a cause of *e*.

Although both approaches alike therefore endorse *c* as a cause, arguably even in this simple case the traditional approach is a little troublesome. It yields a quantitative answer: $\mathrm{p}_{\mathrm{c}}^{\mathrm{ante}}(e) - \mathrm{p}_{\sim\mathrm{c}}^{\mathrm{ante}}(e) = 0.5$, so *c* is deemed a cause of *e* 'to degree 0.5'.[7] At first glance, this might seem appropriate. But, in fact, *do* we judge *c* a cause only 'to degree 0.5'? On the contrary, if the bomb does explode then it seems we more naturally judge *c* just to be the cause of that *simpliciter*, with equal intuitive strength as if the triggering had been entirely determinate—which is just the result that [EP] gives.

[5] Modulo complications concerning ex post counterfactuals, on which see more later.

[6] I intend '$\sim c$', here and elsewhere, as convenient shorthand for specific contrast events, as dictated by context. I intend no commitment to the notion of negative events. Sometimes, exactly which contrast event '$\sim c$' denotes may be unclear. I omit discussion of what to do in those cases because in all examples in this paper, save that discussed explicitly in Sect. 8.1, again the salient locus of philosophical dispute lies elsewhere.

[7] This assumes that a probability-raising criterion such as [EP] is capable of being interpreted quantitatively at all. If $\mathrm{p_c}(e) - \mathrm{p}_{\sim\mathrm{c}}(e) = 0.7$, say, I want to say that this equates to our judgment thinking *c* 'probably' a cause of *e*. Strictly, therefore, our judgment is being deemed sensitive to the *extent* to which $\mathrm{p_c}(e) - \mathrm{p}_{\sim\mathrm{c}}(e) > 0$. In the ex post case, even this will always be 0 or 1.

Of course, notwithstanding its quantitative formula, the standard approach is not committed to analyzing degree of causation. However, although it would take us beyond the scope of this paper to do so, one could argue that there should be continuity between qualitative and quantitative analyses so that not only is a non-zero probability difference necessary for causation, but also the greater that difference the greater the degree of causation. And in so far as causal intuition, slippery enough even in qualitative terms, can also be tracked quantitatively, it then seems that here an ex post approach does it better.

3 A psychological interpretation

As mentioned, $\mathrm{p}_{\mathrm{c}}^{\mathrm{post}}(e)$ always takes the value 1 or 0, depending only on whether *e* actually did occur or not. $\mathrm{p}_{\sim\mathrm{c}}^{\mathrm{post}}(e)$, i.e. the counterfactual, will be 1 or 0 too, modulo complications to be discussed later. It follows that formula [EP], i.e. $\mathrm{p}_{\mathrm{c}}^{\mathrm{post}}(e) - \mathrm{p}_{\sim\mathrm{c}}^{\mathrm{post}}(e)$, will yield integer scores of 1, 0 or -1. In words, and adopting a psychologistic interpretation of [EP], our judgments of causation in singular cases are typically not a matter of degree; rather, they are all or nothing. If [EP] is correct, that is, we deem something either to be a full cause (or hindrance) or else not a cause at all. No scope is left for probabilistic nuance or qualification.

Why do we adopt this attitude? The obvious psychological explanation is that, at least where causation is concerned, we are *natural determinists*.[8] In particular, a commitment to determinism implies we think that, in reality, once *c* occurs the (objective) probability of *e* is already either 0 or 1. Accordingly, our causal judgments track the ex post formulation of [EP], which in return can thus be seen as a way of operationalizing a determinist attitude. True enough, [EP] in itself implies only that we wait until sure of an outcome before attributing causation, and that is not yet equivalent to belief in determinism. Perhaps there are other persuasive motivations for so waiting. But it is hard to imagine *why* we should be so resistant to probabilistic causal attributions—except as a by-product of a belief in determinism.

Notice also that our account is committed to 'counterfactual definiteness'—namely that if *c* had not occurred, then there is a fact of the matter as to whether *e* would or would not have. This corresponds to a Stalnaker-style rather than Lewis-style semantics. Indeed, strictly speaking, the psychological determinism described above really boils down just to this counterfactual definiteness.[9]

According to [EP], whether *c* is a cause of *e* is determined only later in time, after *c* itself has occurred. This kind of temporal extrinsicness is contrary to the standard ex ante view. Given determinism, it is also rather peculiar metaphysically—for on a deterministic view, it should be determinate from the moment *c* occurs whether or not *e* will also occur. Yet while peculiar metaphysically, such extrinsicness is less so psychologically. Perhaps the situation in the eyes of our causal judgment might be

[8] I have in mind a Laplacean view of determinism, according to which the instantaneous state of the world at any time uniquely determines the state at any other time. I will not defend this exact formulation here as, again, the salient locus of philosophical dispute lies elsewhere.

[9] I thank an anonymous referee for emphasizing the points in this paragraph.

described more perspicuously by saying that whether *c* is a cause is sometimes only *revealed* to us after the fact. In our judgment's eyes, in actuality c was already a full cause or not all along (because of determinism), it is just that sometimes we require the subsequent unfolding of events before we can settle our own uncertainty concerning the matter. It is true that on occasion we may feel sure of *e* even before it occurs—but not always. In the general case, therefore, only the ex post formulation tracks our judgment reliably.

Historically, the link between determinism and causation is of course familiar, indeed the latter notion was for a long period almost a synonym for the former. Two problems for causal determinism have motivated the appeal in more recent times to a probabilistic view, but neither problem tells against this paper's approach.

First, according to many at least, modern science suggests strongly that the world itself is *not* deterministic. But nothing here denies the existence of irreducibly indeterministic physical processes. Indeed, Example One was a case that *assumed* such processes and yet nevertheless was easily analyzed by our scheme. The claim is not that the physical universe itself is deterministic, rather only that our causal judgment assumes that it is. Plausibly, perhaps the human brain, when faced with apparent indeterminacy, always explains that as being due merely to ignorance of the underlying deterministic processes, in turn maybe because in everyday life this tactic so often turns out to be productive.[10] The brain may or may not be right to adopt this blanket metaphysical attitude—that is a matter for science. The claim here is merely that it does. (See Sect. 9 for more on the psychological evidence.)

The second problem has been the centrality—indeed seeming indispensability—of probabilistic causation to scientific practice. For example, 'smoking causes cancer' is presumably a valuable scientific discovery, yet the causal claim is obviously probabilistic since smoking does not make cancer certain but rather only more likely. Moreover, this time the problem also threatens to infect the psychological interpretation, because it seems that often our causal judgments are correspondingly probabilistic too. For instance, we easily judge that a particular cancer was 'probably' caused by smoking. I shall explain later though (Sect. 6) how such probabilistic judgments are still compatible with this paper's underlying psychological picture of us as determinists.

We require one more piece of preliminary work. Although a determinist takes the ex post counterfactual $\mathrm{p}^{\mathrm{post}}_{\sim \mathrm{c}}(e)$ always to have a value of either 0 or 1, often, it being counterfactual, we may never know which. Our causal judgment is then forced to appeal instead to its best guess as to whether *e* would occur or not. But given that we are determinists, the uncertainty here is perceived to be merely epistemic. As a result, the 'best' guess is $\mathrm{p}^{\mathrm{post}}_{\sim \mathrm{c}}(e)$'s expected value, conditionalized on our information set restricted to the actual world. What, in turn, is the best guess about this epistemic probability? The answer will depend on the details of the case. Often, it will simply be equal to $\mathrm{p}^{\mathrm{ante}}_{\mathrm{c}}(e)$, i.e. to the *ex ante* value. (See Sect. 6 for

[10] Sometimes I hit a golf ball straight, sometimes crooked. But rather than attribute the outcomes each time to indeterminism, instead I assume always that some aspect of my technique was the (determining) cause. And that assumption is surely not only sensible but also productive, as only in this way am I motivated to search for these causes and thereby to improve.

more on this connection between epistemic and ex ante probability.) But other times, a particular causal model may license a different answer.[11]

For example, let $c =$ John smokes, and $e =$ John develops lung cancer. How should we evaluate $p^{post}_{\sim c}(e)$, i.e. the ex post counterfactual 'if John had not smoked, he would have developed lung cancer'? Suppose that 5% of non-smokers develop lung cancer, while 20% of smokers do. Then the relevant ex ante counterfactual, i.e. $p^{ante}_{c}(e)$, is 0.05—but this may be the wrong evaluation of the *ex post* counterfactual $p^{post}_{\sim c}(e)$. For suppose that there are three kinds of people: type A are very susceptible and get lung cancer regardless of whether or not they smoke; type B are moderately susceptible and get lung cancer if and only if they smoke; and type C will not get lung cancer regardless of whether they smoke. We can infer that 5% of the population are of type A, and 15% of type B. Conditionalized on the actual-world information set that he smoked and developed lung cancer, we can in turn infer that John was either of type A or of type B, but we have no evidence which. Accordingly, the best guess as to $p^{post}_{\sim c}(e)$ is our best guess that John was of type A, i.e. 0.25 (i.e. 5%/20%)—and *not* the ex ante probability 0.05. Other underlying models of lung cancer could generate different answers again.

The claim here is regarding our psychology—that in epistemically imperfect circumstances our causal judgment does not evaluate $p^{post}_{\sim c}(e)$ to be the disjunction 0 or 1, but rather to be a probabilistic best guess.[12] On the psychologistic interpretation, when applying [EP] we must allow in this way for how our judgment incorporates epistemic considerations into its evaluation of $p^{post}_{\sim c}(e)$.[13]

We are now finally ready to turn to [EP]'s chief attraction, namely that it delivers the correct results in key otherwise problematic cases, where 'correct' is read as tracking our causal judgments successfully. There are a plethora of such cases. In this paper I shall consider only a representative few, but analogous solutions apply to others as well.

4 Chance-lowering causes

Example Two: Golf Ball (due originally to Deborah Rosen and much discussed since). A golfer slices her chip shot way to the right, but by good fortune her ball hits a tree and deflects directly into the hole. For $c =$ the golfer slices the ball, and $e =$ the ball goes into the hole, we want to say that c does cause e, via the

[11] Again, I thank an anonymous referee for pointing this out to me, and for the lung cancer example that follows.

[12] This accords with standard decision theory. When planning for the future we rely on our knowledge of what our present actions will cause, and we calculate that in turn via the epistemic probabilities of those actions' effects—even if we think the world is ultimately deterministic.

This also facilitates the continuity mentioned earlier between qualitative and quantitative analyses of causation. In particular, according to [EP] c is deemed a full (zero) cause of e iff $p^{post}_{\sim c}(e)$ equals 0 (1); for all intermediate values of $p^{post}_{\sim c}(e)$, c is a 'partial' cause. But in order to generate and order these intermediate cases, we need to allow our judgment to evaluate $p^{post}_{\sim c}(e)$ probabilistically.

[13] We therefore need to amend the interpretation of the left-hand probability in [EP] as well. In particular, if the right-hand probability is epistemic then, for the sake of the formula's coherence, so should be the left-hand one too. As a result, in [EP] $p^{post}_{c}(e)$ should be interpreted as our judgment's best guess regarding the objective ex post probability, rather than the ex post probability itself. But since, by assumption, whether e occurs is known, the two interpretations yield the same value anyway so this is not a problem.

intermediate event of deflecting off the tree.[14] The problem is that, on the standard ex ante view, this implies a chance-lowering cause:

$p_c^{ante}(e) = 0.001$, say, given that almost all sliced chips end up in the bushes.
$p_{\sim c}^{ante}(e) = 0.05$, say, given that a non-sliced chip hit squarely would be much more likely to end near the hole.
so $p_c^{ante}(e) - p_{\sim c}^{ante}(e) < 0$.

But now we see:

$p_c^{post}(e) = 1$, given that the sliced ball did in fact deflect into the hole.
$p_{\sim c}^{post}(e)$: our judgment has insufficient information to evaluate this ex post counterfactual precisely. As discussed above, in such circumstances it must therefore appeal instead merely to its best guess. Here, this will likely be the ex ante probability, which, by assumption, is 0.05.[15]
so $p_c^{post}(e) - p_{\sim c}^{post}(e) = 1 - 0.05 =$ almost 1.

Thus, by going ex post, *c* is now endorsed as (almost certainly) a cause of *e*, as desired. (More on this example later.)

5 Chance-raising non-causes

Hitchcock (2004) and Schaffer (2000) both rightly emphasize the importance of problem cases 'the other way round', i.e. of chance-raising by apparent non-causes. Here is one that Hitchcock himself selects (2004, p. 410) as being especially intractable.

Example Three: Two Bullets (due originally to James Woodward). Two gunmen fire simultaneously at a vase. Each gunman's bullet has an independent probability 0.5 of hitting (and hence shattering) it. Suppose that, in fact, the first bullet does indeed hit the vase, but that the second one flies wide.

For $c =$ the firing of the second bullet, and $e =$ the shattering of the vase, we want to say that *c* does not cause *e* because the second bullet flew wide. The problem for the standard ex ante view is that this implies that a non-cause is chance-raising:

[14] Or so at least our judgment is traditionally deemed to go here. For those who do not share it, one of the many other framing narratives in the literature may be substituted. What matters for our purposes is only the basic structure of the case, namely that a token effect occurs in an unusual way, leading to a mismatch between attribution of token causation and increase in ex ante probability. (Remember that we are concerned only with physical causal responsibility, not with moral or explanatory responsibility.)

[15] As noted in the previous section, the best guess is not *necessarily* the ex ante probability. It depends on what causal structure we believe to be present, something not fully captured by the quoted probabilities. What matters for our purposes, in both this and other examples, is whichever evaluation of $p_{\sim c}^{post}(e)$ is actually driving our causal judgment, since it is that judgment that ultimately we are seeking to account for. Here, our judgment seems informed by the assumption that not slicing the ball (rather than slicing it) would yield a greater chance of the ball going in the hole, and any evaluation satisfying that qualitative constraint leads to [EP] delivering the desired verdict. (The ex ante probability is only one such evaluation.) More outré possibilities, wherein not slicing the ball would *not* have altered the chances of the ball going into the hole, do not seem to be informing our judgment here. But if they were, so that whether we sliced the ball made no difference, arguably we would no longer judge slicing the ball to be a cause of it going in the hole—and [EP] would in turn now reflect that. (In Sect. 8.3, I analyze an example in which [EP] tracks in this way how our judgment varies with background causal assumptions.)

$$p_c^{ante}(e) = 0.75,\ p_{\sim c}^{ante}(e) = 0.5,\ \text{so}\ p_c^{ante}(e) - p_{\sim c}^{ante}(e) > 0.$$

But now analyze in terms of ex post probabilities: first, clearly $p_c^{post}(e) = 1$. Given that the first bullet is stipulated to hit the vase independently of the second bullet, we also know that the ex post counterfactual $p_{\sim c}^{post}(e) = 1$ too. Intuitively, whether the second bullet was fired made no difference to the fate of the first bullet. (I discuss the evaluation of such ex post counterfactuals in more detail in Sect. 7.) So $p_c^{post}(e) = p_{\sim c}^{post}(e)$, which is exactly what we want when c is a non-cause. Thus the ex post approach again tracks our causal judgment accurately.

Moreover, recapitulating a point from Sect. 2, consider d = the firing of the first bullet. It is readily seen that (in obvious notation):

Ex ante: $p_d^{ante}(e) - p_{\sim d}^{ante}(e) = 0.75 - 0.5 = 0.25$.
Ex post: $p_d^{post}(e) - p_{\sim d}^{post}(e) = 1 - 0 = 1$.

Both approaches yield that d is a cause of e, as desired. But only by going ex post is the first bullet given full causal credit for the shattering of the vase, which I think matches our judgment here much better than a credit merely 'to degree 0.25'.

6 Type and epistemic probabilities

There is an obvious intimate connection between *ex ante* probability and *type* probability. Recall the indeterministic bomb from Example One: the atom decays, thereby triggering the bomb, with probability 0.5. In a population of such bombs, the type probability of a bomb exploding is clearly equal to the ex ante probability of 0.5 (assuming that each device is independent).

As is well known, the type/ex ante probability is in turn central for prediction and other instrumental purposes. For example, the total number of expected explosions would be 0.5 times the number of devices. Often it is the type probability that is deemed relevant to our assignment of moral or legal responsibility, just because it exactly captures what could reasonably have been expected.[16] Moreover, often it is the type probability that is relevant to the practice of science. For example, the claim 'smoking causes cancer' is type-probabilistic. Controlled experiments detect causal relations in proportion to type probabilities. Causal inference from statistics is also on the basis of type probabilities—tracking, say, the correlation between the placing

[16] Indeed, turning this around, Amit Pundik has suggested to me that the real mystery in legal practice is rather why anything *other* than ex ante probability is considered relevant to determining liability. In particular, often legal weight is put not just on the level of risk but also on whether that risk is actually instantiated. For example, drink-driving and having an accident is generally more punished than drink-driving and not having one. This is so even when the ex ante probability of the drunkenness causing an accident was equal in the two cases, indeed even if the ex ante probability in the accident case was actually *lower*. The account of this paper now offers the beginnings of an explanation for this—namely, that our causal judgment naturally seizes on ex post outcomes. When apportioning legal responsibility, it seems there is a mixture of 'reasonable' moral consideration of ex ante chances, plus causal consideration only of ex post outcomes. Hence some legal weight ends up being put on those ex post outcomes, regardless of how 'unfair' it may seem to do so. (Perhaps this point may also shed light on issues surrounding moral luck more generally.)

of atoms in bomb devices and subsequent bomb explosions, we would find that the latter followed the former half the time on average.

Thus ex ante probabilities, through their connection to type probabilities, are clearly ubiquitously useful. Nothing in this paper should be taken as disputing that. But, recall, our account of token causal judgment seemed to leave no role for ex ante probability at all. So just what, on our account, *is* its role? To answer that, it is instructive to compare Examples Two and Three from earlier, and in particular to compare their treatments of the ex post counterfactual $p^{post}_{\sim c}(e)$.

First, in Two Bullets (Example Three), $p^{post}_{\sim c}(e)$ corresponded to the probability that the first bullet would hit the vase given that the second had not been fired. What we needed to know, on our scheme, was this probability's ex post value. In other words, *would* the first bullet have hit the vase if the second hadn't been fired, yes or no? As it happened, our judgment had sufficient information to assess this with certainty. In particular, the crucial fact was that in actuality the first bullet did hit the vase—*independently* of the second having been fired.

Second, contrast this with Golf Ball (Example Two). There, $p^{post}_{\sim c}(e)$ corresponded to the probability that the ball would have gone into the hole had the golfer hit it squarely. It was stipulated that this probability, understood in the ex ante or type sense, is 0.05. But what is its ex post value? That is, *would* the ball have gone into the hole had the golfer hit it squarely, yes or no? Unfortunately though, in the actual world the ball was *not* hit squarely, so this squarely-hit chip shot never existed, and we have no way of saying with confidence whether it would have gone into the hole or not. Therefore, unlike in Two Bullets, our judgment is unable to 'fill in' an ex post probability and so was forced instead to take its best guess at it—which will, depending on the details of the case, often be the ex ante one.

What is the lesson of these two examples? My answer: that our causal judgment fills in an ex post value for the relevant probabilities *whenever it can*. It only resorts to its best guess, and hence to one not valued at 0 or 1, when it has insufficient information to fill in an ex post value. That is the salient difference between the two examples—in Two Bullets we feel we do have sufficient information to evaluate the counterfactual ex post probability, but in Golf Ball we do not. In other words, the binding constraint is *epistemic*. On this paper's account, this is the only route by which ex ante probabilities get to make any appearance at all in our causal judgments. In particular, we (sometimes) resort to them only when the relevant ex post probability is unknown. As far as our causal judgment is concerned, ex ante probabilities are merely instances of epistemic probabilities.

This makes it clear how even a determinist may nonetheless often assign causal responsibility probabilistically. For example, suppose we ask whether smoking was responsible for a particular smoker's lung cancer. The answer will depend on our evaluation of the relevant ex post counterfactual, namely—*would* the cancer have occurred without the smoking, yes or no? Without knowledge of the detailed physical history of the inside of the smoker's lung, we can only make our probabilistic best guess given relevant causal knowledge and statistical data. (As explained in Sect. 3, this best guess will often but not always be the ex ante probability.) Thus, even while believing that in reality the smoking either did or didn't cause the cancer, epistemic constraint leads us to apportion the responsibility only probabilistically.

For its part, meanwhile, a type probability here can be glossed as us being unsure of which token case applies. In particular, invocation of a type probability in a token case can always be viewed as a symptom of some epistemic constraint. By contrast, in type cases the invocation of ex ante rather than ex post probabilities is normal. Overall, we go ex ante in *both* type and epistemic cases because these two are the *same* kinds of case in the eyes of our deterministically inclined causal judgment.

Causation itself is taken by us to be bivalent—something is either a cause or it isn't. Non-trivial probabilities only crop up in type cases and some token cases, each time the source of the uncertainty being epistemic. These probabilities are epistemic in the traditional subjectivist sense of Laplace, i.e. they are just a means of representing our uncertainty concerning what we take to be in actuality a determinate fact of the matter. Thus any probabilistic causal claim is best glossed as elliptical either for an average over a particular population of token cases, or else for uncertainty regarding which token case applies. 'Smoking causes cancer', for instance, should thus be seen as shorthand for 'smoking *sometimes* causes cancer.'[17]

7 Ex post counterfactuals

How do we evaluate ex post counterfactuals such as $\mathrm{p}^{\mathrm{post}}_{\sim \mathrm{c}}(e)$, central to our account throughout? Consider Two Bullets (Example Three): there, for $e =$ the vase shatters, and $c =$ the second bullet fires, we wanted to know whether e would still occur even if c did not. Intuitively, our reasoning seemed to be that since in actuality the vase was shattered by the first bullet, and since the first bullet's flight was independent of the second's, so that shattering is deemed still to happen in the nearest possible world in which the second bullet is not fired. Formally, the nearest $\sim c\&e$-world is deemed closer than any $\sim c\&\sim e$-world. That is, the similarity relation between worlds must endorse the closeness of a possible world in which the vase still shatters even though that shattering occurred *after* the 'miracle' of the second bullet not firing, and even though ex ante the shattering was chancy. This is just the direct analogue of ex post probabilities in the actual world, which likewise assume the occurrence of some post-c events that are ex ante chancy.[18]

[17] The general relation between type and token causal claims is a deep issue (Hitchcock 1995; Eells 1991), which I do not cover fully here. Paradoxes in the literature on probabilistic causality often turn on a mismatch between type probabilities and actual token outcomes, e.g. an effect occurring in an unlikely way as in Golf Ball. This has prompted the suggestion that type and token causation are two entirely distinct things, and that a simple probability-raising criterion is appropriate only to the type case (Sober 1985; Eells 1991). But, in common with most of the literature, I do not support this suggestion. First, I think it is possible to understand the two in a unified scheme, as this paper explains. Second, other kinds of problem for a probability-raising account, e.g. overdetermination cases, arise in both type and token contexts alike.

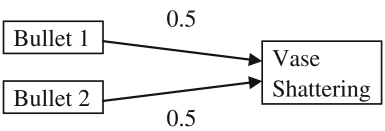
[18] On this point, as well as on others, the account here differs from those few previous ones that have sought to analyze probabilistic causation in a deterministic way (Barker 2004; Ramachandran 1997—although Ramachandran has since endorsed a different view). Those other accounts invoke a form of deterministic counterfactual dependence, but not with the counterfactuals being evaluated in the ex post way just described.

In particular, this example suggests that when assessing the nearest possible $\sim c$-world, we hold fixed all events that are not causally downstream of c, regardless of whether those events are themselves ex ante chancy.[19] In this case, that means holding fixed that the first bullet hits the vase, since it is stipulated that the first bullet's fate is not causally influenced by the second's. By modus tollens, perhaps the best argument for this way of evaluating counterfactuals is precisely our initial judgment that the second bullet was not the cause of the vase's shattering—for if, contrary to our theory, we *could* have thrown the vase's shattering into question by not firing the second bullet, then I think we *would* deem its actual firing a (possible) cause of the shattering after all.

Of course, the above raises the issue of circularity. We are using these counterfactuals to explicate the notion of causation, and yet now explicit reference to a causal notion is made when evaluating them. It follows that formula [EP] therefore cannot be a reductive definition. But it currently seems dubious that any reductive semantics for causal counterfactuals is available anyway (Woodward 2003; Schaffer 2004). [Cartwright (1979) and Eells (1991) argue against any reduction of causation to probabilities; Kvart (1986) argues against a non-causal analysis of counterfactuals, but maintains a reductive probabilistic theory of causation.] Further, once we are willing to evaluate such counterfactuals non-reductively, one attractive alternative that thereby becomes available is to evaluate them via explicit causal models instead (Pearl 2000; Spirtes et al. 2000). Arguably, this technique offers many advantages in clarity over the alternative of Lewis-style nomological comparisons of similarity (Woodward 2003, pp. 133–145).

Returning to the Two Bullets example for illustration, the causal model of that case can be represented graphically.

Two Bullets

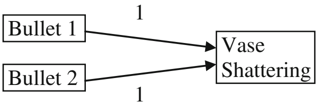


[19] This approach to evaluating indeterministic counterfactuals is endorsed widely, for instance by Edgington (2004) and Schaffer (2004). [Although see Maudlin (2007) for an opposing view.] An example of Edgington's (due originally to David Johnson) captures the essential point—suppose I am watching a distant indeterministic lottery on television, and a relative of mine is drawn to be the winner. Now consider, what if while watching the draw I had scratched my nose? We have a strong intuition that this scratching would have made no difference to the draw's outcome, precisely for the reason that the draw could not have been affected causally by it. But consider how our intuition changes if the counterfactual scenario is instead that the lottery wheel was spun in a different way. Now of course it seems much more questionable whether the outcome of the draw would have remained the same, precisely because the draw's outcome would now indeed have been causally downstream of the counterfactual's antecedent.

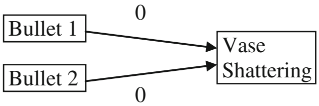
The counterfactual of the second bullet not firing can then be readily evaluated by modeling it as an *intervention* on that node, canceling its input, roughly speaking, and the impact of that intervention then traced on the graph.[20]

The issue here then becomes how to adapt this technique to our own ex post approach. The clue lies in our causal judgment's determinism. In its eyes, the true causal model cannot be the probabilistic one of the above diagram, but rather must correspond to one of the following four possibilities.

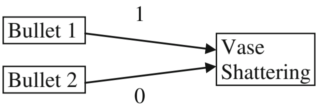
Scenario (a)



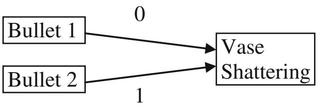
Scenario (b)



Scenario (c)



Scenario (d)



The actual course of events, in which the first bullet hit the vase but the second missed, establishes to our causal judgment that it is scenario (c) that is the true one. Once we have thus settled conclusively on which causal structure actually obtains, we can then use that structure to evaluate the counterfactual of not firing the second bullet. Clearly, the verdict is that the vase would still shatter, and thus $\mathrm{p}^{\mathrm{post}}_{\sim\mathrm{c}}(e)$ is evaluated easily.

8 Further problem cases

A major benefit of [EP] is that it also sheds new light on several other important issues and problem cases in the literature.

8.1 Contrast classes

Recall Example Two: a golfer slices her chip shot way to the right (c), but by good fortune her ball hits a tree and deflects directly into the hole (e). c causes e, yet as we saw earlier, $\mathrm{p}^{\mathrm{ante}}_{\mathrm{c}}(e) - \mathrm{p}^{\mathrm{ante}}_{\sim\mathrm{c}}(e) = 0.001 - 0.05 < 0$.

[20] Note that therefore the circularity involved here is not vicious (Woodward 2003). In particular, the assumed causal notion of an intervention invokes the causal relation between an exogenous variable and the bullet firing, and *not* the one between that firing and the vase shattering.

As well as being an example of a chance-lowering cause, this also raises a further issue. In particular, one especially interesting analysis, favored by Hitchcock (2004, p. 405), appeals to different contrast classes. Here, in context '$\sim c$' is ambiguous between different contrast events—for instance between the golfer hitting the ball squarely, or her swinging and missing it altogether. Given the latter choice, plausibly $p_{c}^{ante}(e)$ will again be greater than $p_{\sim c}^{ante}(e)$, since $p_{\sim c}^{ante}(e)$ is now presumably virtually zero. Thus, Hitchcock claims, all depends on which contrast we choose. We can accept as true both the claim that 'the golfer's slice *caused* the hole-in-one', and also the claim that 'the ball landed in the cup *despite* the badly sliced shot'—because 'we can hear them … as being made relevant to different alternatives' (p. 405).

Although I do not think this solution is quite right, nevertheless I agree with much of it. '$\sim c$' indeed needs to be disambiguated, and different disambiguations indeed result in different values for $p_{\sim c}^{ante}(e)$.[21] Moreover, the above contrastive account of Golf Ball seems to me to be entirely accurate with regard to the ex ante/type probabilities. Thus as a matter of prediction or advice, for instance, endorsement of slicing the ball does indeed depend critically on whether the contrast is with hitting it squarely or with not hitting it at all.

Recall next our own previous solution to Golf Ball: $p_{c}^{post}(e) - p_{\sim c}^{post}(e) = 1 - 0.05 > 0$. Why should we prefer this solution to Hitchcock's? Here are two reasons. First, even for $\sim c =$ hitting the ball squarely, still our judgment remains clearly that c is a token cause of e—contrary to the contrastive ex ante account. Intuitively the reason is that, however fortuitously, *in this case* slicing the ball meant it going into the hole via the deflection whereas hitting the ball squarely would probably only have left it near the hole, not in it. This is exactly the reasoning captured by our ex post formula. Unlike in the type case, we hear no real ambiguity regarding the token claim. Rather, I think c is endorsed as a token cause so firmly precisely because all disambiguations of $\sim c$ that are plausible in the context yield alike some very small $p_{\sim c}^{post}(e)$ or other.

Second, for $\sim c =$ missing the ball altogether, the relevant increase in ex ante probability associated with c is minuscule: on the (already generous) figures above, from 0 to 0.001. As with earlier examples, I think our judgment endorsing c as a token cause is rather more emphatic than that, and only the ex post approach captures this.

8.2 Probability pooling

Sometimes, ex post probabilities may be difficult to assign because the effects of different causal contributions are inseparable.

Example Four: Rope (due to Christopher Hitchcock). If Gene or Pat tugs on a rope individually, the rope will break with probability 0.5. If they tug on it together, it will break with probability 0.75. On a given occasion, they tug together and the rope indeed breaks. Was Gene's tug a cause of the rope breaking?

[21] See, e.g., Schaffer (2005) and Northcott (2008) for arguments in favor of a contrastive theory of causation generally.

Let c = Gene's tug, $\sim c$ = Gene's standing idly by (i.e. only Pat tugs), e = the rope breaks. Clearly, $p_c^{post}(e) = 1$. But $p_{\sim c}^{post}(e)$ is hard to evaluate due to insufficient information—if Gene had not tugged, would the rope still have broken? Well, maybe. Accordingly, all we feel able to say is that Gene's tug 'might' have been the cause. Our hesitant response is, on this paper's account, a direct consequence of the difficulty in evaluating $p_{\sim c}^{post}(e)$. Without more details about the causal processes that underlie the rope-breaking, the counterfactual is 'uncertain'.[22]

Compare now Examples Three and Four, namely Two Bullets and Rope. A particularly telling detail is that so far as *ex ante* probability is concerned the structures of the two cases are identical—and yet our reactions to them are nevertheless very different. So there must be more things in heaven and earth, causally, than are dreamt of in ex ante probability. Formally, the only difference between the two cases is that whereas the effects of the two bullets are clearly separable, those of the two rope tugs are not. As a result, the relevant ex post counterfactual is evaluable only in Two Bullets. Our judgment tracks exactly this difference, suggesting again that the probability to which it is sensitive is ex post, not ex ante.

It also suggests that our judgment does *not* treat causation as mere contribution to a probability 'pool'.[23] On the contrary, causal contributions are pooled only ever as a second-best formal maneuver, when epistemic constraint forces that upon us. In our hearts, so to speak, we are determinists and feel that no probability is ever really involved, pooled or otherwise. Whenever it can, as in Two Bullets, our judgment therefore always seeks to avoid pooling and instead to assign all-or-nothing responsibility to individual causes.

8.3 When details matter

Sometimes, we may have more than one intuition about the value of $p_{\sim c}^{post}(e)$. In such cases, as noted earlier, I think our causal judgment depends on the empirical details, and in a way well tracked by formula [EP].

Example Five: Weed (adapted from one due originally to Nancy Cartwright). A weed is sprayed with defoliant (c) but nonetheless survives (e). Assume that $p_c^{ante}(e) = 0.3$, and $p_{\sim c}^{ante}(e) = 0.7$. That is, ex ante, spraying the defoliant typically reduces the weed's chance of survival. But because, in this particular case, in fact the weed does survive despite the spraying, so $p_c^{post}(e) = 1$. What is $p_{\sim c}^{post}(e)$? That is, *would* the weed have survived if it had not been sprayed, yes or no? There seem to be two possible answers, depending on the biochemical details.

[22] Paul Noordhof (2004, p. 190) raises an objection to deterministic accounts of probabilistic causation, namely that while they assume that $p_c^{ante}(e)$ is neither 0 nor 1 (due to the case being probabilistic), at the same time they also happily assume that $p_{\sim c}^{ante}(e) = 0$. But what justifies such an asymmetric treatment of $p_c^{ante}(e)$ and $p_{\sim c}^{ante}(e)$? However, this perceptive criticism [aimed at Ramachandran (1997) and Barker (2004)] does no damage to the account in this paper. Our ex post probabilities are evaluated as 0 or 1 given c or $\sim c$ alike. The only exception is for some $\sim c$ cases and, as we have seen, for those there is a ready *epistemic* explanation—namely, that in such cases (e.g. pooling ones like Rope) the exact value of the counterfactual ex post probability is unknown.

[23] In this respect, as in others, our account diverges from those of Lewis (1986) and Humphreys (1989). See Hitchcock (2004) for more discussion of pooling.

First, we might evaluate $p^{post}_{\sim c}(e) = 1$, perhaps on the grounds that this weed was revealed by its survival of the spraying therefore to be in the hardiest 30% of its kind. Thus for sure it would also have survived if there had been no spraying. This assumption would be justified if, for instance, knowledge of the biochemistry showed that the action of the defoliant was causally independent of the normal risks undergone by the weed. That is, the defoliant 'added' an extra 0.4 risk in addition to the baseline 0.3 risk run by the weed anyway. Thus even if the defoliant had never been sprayed, still our judgment holds fixed the weed's survival of the separate 0.3 risk, since (by assumption now) that latter risk is not causally downstream of the spraying. (In exactly the same way, in Two Bullets we held fixed the outcome of the first bullet when assessing the counterfactual of not firing the second.) Under such assumptions, here we get $p^{post}_{c}(e) - p^{post}_{\sim c}(e) = 1 - 1 = 0$. So the spraying is not endorsed as a cause of the weed's survival, just as the gunman's second bullet was not endorsed as a cause of the vase shattering.

The second possibility is that we instead evaluate the ex post $p^{post}_{\sim c}(e)$ to be the ex ante $p^{ante}_{\sim c}(e)$, i.e. 0.7. To illustrate the case for this alternative, suppose the biochemical details this time are such that the action of the defoliant in some way blocks out the weed's normal 0.3 risk process. As it turns out, the weed survives. But if the defoliant had never been sprayed, would it still have done so? Since we are now assuming that the normal risk process *was* causally downstream of the spraying, when evaluating the counterfactual we may no longer hold that process's outcome fixed. Thus we cannot rule out the possibility this time that, had the defoliant not been sprayed, the weed would have died. Rather, as often in situations of epistemic constraint, we are instead forced to revert back to the ex ante probability as our best available proxy for the ex post one.[24]

Formally, the second case therefore yields: $p^{post}_{c}(e) - p^{post}_{\sim c}(e) = 1 - 0.7 > 0$. In words, now the spraying is endorsed as a (possible) cause of the weed's *survival*. Is this conclusion defensible? I think so. Consider what we are assuming in this scenario: that spraying the defoliant interfered with the weed's normal risk processes in such a way that the weed survived. That is, spraying the defoliant in this case definitely led the weed to live, but had we *not* sprayed it then the weed might *not* have lived. Given *those* assumptions, it seems clear that the spraying indeed left the weed better off—in *this* case. Intuition is no longer repelled. Of course, for *general* advice we should turn instead to the type probability, which would endorse a spray-to-kill strategy. But for this particular token case, given the assumptions behind this particular evaluation of $p^{post}_{\sim c}(e)$, not so.[25]

[24] No doubt other assumptions regarding the biochemical details could yield yet other values for $p^{post}_{\sim c}(e)$, and thus corresponding variation in the causal verdicts regarding *c*.

[25] Eells (1991, p. 290) and Hitchcock (1995, p. 269) each notes a puzzling apparent asymmetry between Examples Two and Five, i.e. between Golf Ball and Weed. Both times, *c* lowers the ex ante probability of *e*. Yet only in Golf Ball does intuition nonetheless endorse *c* as a token cause. That is, intuition endorses the slice as a cause of the ball going into the hole, while by contrast it seems to rebel at endorsing the spraying of defoliant as a cause of the weed's survival. But, as explained in the text, intuition in the latter case varies subtly with type or token context, and also with our assumptions regarding the biochemical details and thus evaluation of $p^{post}_{\sim c}(e)$. Once these important extra matters are clarified, the asymmetry in intuitive response melts away.

9 Causation versus causal judgment

If the account of this paper is accepted, and if furthermore it is interpreted as a thesis about causal judgment, then we are left with two options regarding its implications for causation itself:

(1) Causation itself is not deterministic; it is merely that our causal judgment mistakenly assumes so.
(2) Causation really is deterministic.

Classical Humean projectivists, who presumably identify causation with causal judgment, are thereby committed to option 2—and thus now to the view that probabilistic causation does not exist, properly speaking, rather only epistemic uncertainty. Other metaphysical positions, by contrast, are free to countenance the separation of causation and causal judgment posited by option 1. In itself, this paper's formalism is compatible with either metaphysical choice, but in the remainder I shall focus on option 1. One motivation is that, as we shall see in a moment, this yields us a new defense of causation as ex ante probability-raising.

I take it that ultimately the project here is metaphysical, i.e. a concern with causation's place in the structure of the world. The way that philosophers have typically investigated this is via perusal of thought-examples; yet the data, so to speak, yielded by such perusals can only be particular causal judgments. To treat those judgments uncritically as true metaphysical guides is to assume that they track, always and everywhere, causation itself. It might be argued that we have no choice but to assume that—after all, the thought runs, how else could we proceed? But this paper's account suggests, on the contrary, that sometimes we may have good reason to *expect* our causal judgment to diverge radically from causation itself. In particular, if humans are indeed natural determinists, then in certain probabilistic cases this psychological bias will lead to predictable distortions. That in turn gives us the license, indeed obligation, to treat our judgments in such cases critically. (By way of analogy, our well-founded understanding of the human visual system licenses us, indeed obliges us, to disregard certain optical perceptions as explicable illusions.)

A novel method for defending the ex ante probability-raising criterion of causation thereby becomes available. In particular, apparently damning counterexamples (or rather our judgments in them) can now be disregarded as mere artifacts of a naively determinist causal psychology. Of course, not just any explaining-away of counterexamples is legitimate; such maneuvers must also be constrained in a principled manner. That is why it is necessary to state the psychological theory precisely, as per [EP]. Further, the underlying psychological hypothesis should be independently plausible. That is why it is significant that the pattern of ex post causal judgment represented in [EP] can be seen as a natural consequence of a determinist attitude, since our having that attitude in turn carries independent plausibility (Sect. 3).

Any such psychological hypothesis is of course ultimately an empirical one. We thus open a welcome new route for empirical input into metaphysical inquiry. Much 'experimental philosophy' so far has tested whether certain intuitions claimed to be widespread are indeed so, or has tracked what factors those intuitions are sensitive to. That is different from what I am proposing here, which is that findings from psychology can license us to *disregard* certain intuitions altogether. Empirical evidence thus bears directly on inquiry into the underlying metaphysics, not just psychology. A priori discussion can be—and ought to be—augmented in a principled way.

The only empirical evidence that this paper provides is informal, namely our intuitive reactions to certain thought-examples. What does the extensive causal psychology literature say on the matter? To my knowledge, as yet nothing directly—let us see why. At issue is whether we judge causation probabilistically ex ante or deterministically ex post. It might be thought that decisive results do exist already and that they tell against this paper's position. There is ample experimental evidence that even infants, for instance, naturally interpret causation probabilistically. But care must be taken when assessing such evidence's implications. In particular, what has actually been shown is that infants can infer causal relations from probabilistic data—and that is quite consistent with the view of this paper. For instance, in Example One we might infer the probabilistic causal relation that half the time placing the atom in the box causes the bomb to explode. But as explained earlier (Sect. 6), such a type-inference is readily compatible with our account. The understandable overall emphasis in the cognitive psychology literature on type-level causal learning means that its empirical results are not decisive with regard to our particular debate. Moreover, even in token cases, assigning retrospective causal responsibility probabilistically is similarly not decisive, so long as examples of such a practice can be explained each time as an instance of epistemic constraint forcing us to evaluate the relevant ex post counterfactual via a probabilistic best guess.

Generally, precisely because our account is consistent with the ubiquity of probabilistic causal judgment, therefore one has to work hard to engineer critical cases where the judgments predicted by this paper's thesis actually do diverge from those predicted by a standard probabilistic view. But, notice, we have *already encountered* a number of just such cases—they are precisely those discussed in Sects. 4, 5 and 8. And our judgments there tracked the ex post formula [EP], not standard ex ante probability-raising. In lieu of future empirical investigation, those judgments are therefore important prima facie evidence in our account's favor.

To sum up: the account of causal *metaphysics* as ex ante probability-raising can be saved from particular counterexamples by conjoining it with an account of our causal *psychology* as determinist. Empirical evidence regarding the latter can thus serve to save (or undermine) the former. As mentioned in Sect. 1, such a salvation is desirable for independent reasons; a new path to it is therefore good news.

Methodologically, I have assumed that our access to causation is through the filter of our causal judgment. Therefore any account of causation must implicitly be a composite account of both our judgment and what we are judging, i.e. of both psychology and metaphysics. It seems dubious to suppose we can know a priori any particular causal judgment to be infallible. Thus we need instead to formulate different metaphysics/psychology packages, and then to test each package as a whole against the 'data' of our experience of the world, including especially our causal judgments (understood broadly to include also their implicit reflection in the practice of science). This paper's is just one possible such package. Of course, there may be others that fit the data too. But a package that seems clearly *not* to, is the one that conjoins ex ante probability-raising with the simple psychology that all causal judgments are veridical. Thus the necessity of adjusting the second component in order to save the first.

Finally, it must be conceded squarely, there remains one important class of cases in which this paper's formal account does still falter, namely cases of causal overdetermination (including pre-emption). That is, in psychologistic terms, the particular psychological hypothesis that we are natural determinists does *not* explain away the discrepancy between a token counterfactual account of probability-raising (such as ours) on one hand, and our causal judgments in overdetermination cases on the other. In this respect, [EP] exhibits the usual vulnerability of a simple counterfactual account. Perhaps this paper sheds new light even on this well known challenge though, albeit only indirectly. In particular, I believe that the same general methodological strategy introduced here, applying psychology to metaphysics, may prove effective in overdetermination cases too. But demonstration of that will have to wait for another day.

Acknowledgements I would like to thank the following for helpful comments: Carl Craver, Eric Hiddleston, Christopher Hitchcock, Boris Kment, Monika Koeppl, and Kent Staley. I would especially like to thank an anonymous referee for this journal. Ancestor versions of this paper benefited from being presented at: the Society for Exact Philosophy conference at the University of California San Diego in May 2006; the Causality and Probability in the Sciences conference at the University of Kent in June 2006; and a Philosophy, Neuroscience and Psychology seminar at Washington University in St Louis in September 2006.

References

Barker, S. (2004). Analysing chancy causation without appeal to chance-raising. In P. Dowe & P. Noordhof (Eds.), *Cause and chance* (pp. 120–137). London and New York: Routledge.

Cartwright, N. (1979). Causal laws and effective strategies. *Noûs, 13*, 419–437.

Dowe, P. (2004). Chance-lowering causes. In P. Dowe & P. Noordhof (Eds.), *Cause and chance* (pp. 28–38). London and New York: Routledge.

Edgington, D. (2004). Counterfactuals and the benefit of hindsight. In P. Dowe & P. Noordhof (Eds.), *Cause and chance* (pp. 12–27). London and New York: Routledge.

Eells, E. (1991). *Probabilistic causality*. Cambridge: Cambridge University Press.

Hitchcock, C. (1995). The mishap at Reichenbach Fall: Singular versus general causation. *Philosophical Studies, 78*, 257–291.

Hitchcock, C. (2004). Do all and only causes raise the probabilities of effects? In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 403–418). Cambridge, Massachusetts: MIT Press.

Humphreys, P. (1989). *The chances of explanation*. Princeton, New Jersey: Princeton University Press.

Kvart, I. (1986). *A theory of counterfactuals*. Indianapolis: Hackett Publishing.

Kvart, I. 2004. Probabilistic cause, edge conditions, late preemption and discrete cases. In P. Dowe & P. Noordhof (Eds.), *Cause and chance* (pp. 163–187). London and New York: Routledge.

Lewis, D. (1986). *Philosophical papers* (Vol. II). Oxford: Oxford University Press.

Maudlin, T. (2007). *The metaphysics within physics*. Oxford: Oxford University Press.

Noordhof, P. (2004). Prospects for a counterfactual theory of causation. In P. Dowe & P. Noordhof (Eds.), *Cause and chance* (pp. 188–201). London and New York: Routledge.

Northcott, R. (2008). Causation and contrast classes. *Philosophical Studies, 39*, 111–123.

Pearl, J. (2000). *Causality*. New York: Cambridge University Press.

Ramachandran, M. (1997). A counterfactual analysis of causation. *Mind, 106*, 263–277.

Schaffer, J. (2000). Overlappings: Probability-raising without causation. *Australasian Journal of Philosophy, 78*, 40–46.

Schaffer, J. (2004). Counterfactuals causal independence and conceptual circularity. *Analysis, 64*, 299–309.

Schaffer, J. (2005). Contrastive causation. *Philosophical Review, 114*, 297–328.

Sober, E. (1985). Two concepts of cause. In P. D. Asquith & P. Kitcher (Eds.), *PSA 1984* (Vol. 2, pp. 405–424). East Lansing, Michigan: Philosophy of Science Association.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge, Massachusetts: MIT Press.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.