White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/201976/

Version: Accepted Version

Book Section:
Bebb, J. and Beebee, H. (2024) Causal selection and egalitarianism. In: Knobe, J. and Nichols, S., (eds.) Oxford Studies in Experimental Philosophy. Oxford Studies in Experimental Philosophy, 5. Oxford University Press, Oxford, UK, pp. 401-433. ISBN: 9780198918875.

https://doi.org/10.1093/9780198918905.003.0016

This item is protected by copyright. This is an author produced version of a chapter published in Oxford Studies in Experimental Philosophy Volume 5. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Causal selection and egalitarianism[1]

Jon Bebb & Helen Beebee

Please do not cite this version. The published version is: 'Causal selection and egalitarianism' in J. Knobe and S. Nichols, eds., *Oxford Studies in Experimental Philosophy, Vol. 5* (Oxford: Oxford University Press), forthcoming (2023 or 2024).

1. Introduction

Much of the literature on the metaphysics of causation operates with an 'egalitarian' concept of causation. If causation is a matter of counterfactual dependence, then *any* event (or action or fact or state of affairs or whatever) on which the effect counterfactually depends is, equally, among its causes. The sending of the email was caused not only by its author hitting the 'send' button but also by the flow of electricity to the laptop, the presence of oxygen in the room, the Big Bang, the absence of a hungry lion (assuming, as we do in this paper, that absences can be causes), and so on. Similarly if causation is a matter of the manifestation of some power, the transfer of energy, or whatever. While there are differences between the various causes—some may be actions intended to bring about the effect and some not, some entirely routine and expected and some not, some more distant in time and space than others, and so on—those differences do not amount to any difference in their status *as* causes. And most of the relevant theorists implicitly or explicitly assume that this egalitarian concept is *our ordinary* concept of causation: the one we deploy as we go about our daily lives asserting and denying causal claims. We'll call this view—that our ordinary concept of causation is egalitarian—*egalitarianism*.[2]

On the other hand there has been a good deal of recent work in experimental philosophy concerning the phenomenon of 'causal selection': we tend to single out one candidate cause rather than another even when the effect manifestly counterfactually depends on both of them. Causal selection puts pressure on egalitarianism, since it appears to show that we do *not*, in general, regard all of the candidate causes—those events that satisfy some

[1] Thanks to the audience at Kristie Miller's *Time and Causation* workshop at the University of Sydney and to Wesley Buckwalter and Ann Whittle for helpful comments and suggestions. Helen Beebee is grateful for the support of the Engineering and Physical Sciences Research Council in writing this chapter (grant no. EP/W01081X/1).

[2] We have borrowed the term 'egalitarian' from Clarke *et al.* 2015; see also Livengood and Sytsma (2020: 65).

metaphysical conditions on causation, such as counterfactual dependence or energy transfer or whatever—as genuine causes.

At first sight, one might think that causal selection does more than merely put pressure on egalitarianism; rather, it settles the matter in favour of the inegalitarian. This would be considerably too hasty, however. In this chapter we investigate how much pressure the burgeoning x-phi literature on causal selection does in fact put on egalitarianism.

X-phi experiments in this area normally present participants with a vignette wherein some situation or process involving two or more candidate causes of some effect *y* is described—where the relevant actions, events or absences count as candidate causes by virtue of *y*'s counterfactual dependence on each of them. Participants are then normally asked to express their degree of agreement or disagreement with some statement of the form '*x* caused *y*', or '*x* was a cause of *y*', or '*x* was the cause of *y*', or whatever. And it is generally found that participants assign some candidate causes considerably higher ratings than others. Hence causal selection: some candidate causes are 'selected' while others are not. The main focus of work in this area has been to try to figure out the nature of the psychological mechanism (or perhaps mechanisms) that explains this differential treatment of candidate causes: on what basis (or bases) one candidate cause rather than another is selected.

In the face of the now well established phenomenon of causal selection, the egalitarian has two basic moves available to her. One is to grant that an expression of agreement or disagreement with a claim of the form '*x* caused *y*' (or similar) straightforwardly indicates a judgement about whether or not *x* was indeed a cause of *y*, but to argue that the psychological mechanism responsible for that judgement—insofar as it is selective—is defective in some way. That is, rather than manifesting competence with respect to the concept of causation, the mechanism delivers false judgements in those cases where the judgement is that some event on which *y* counterfactually depends is *not* a cause of *y*. In the current dialectical situation one cannot, of course, legitimately claim that the mechanism is defective merely on the basis of the fact that the ordinary concept of causation is egalitarian and our judgements are often out of kilter with this fact; that would be to beg the question against the inegalitarian. But, as we'll see, other reasons can be given, and have been given, for claiming that the relevant mechanism is defective in the sense just described. We'll call this egalitarian strategy *Defective Mechanism.*

The other move is to appeal to broadly pragmatic phenomena, such as pragmatic presupposition or conversational implicature. According to this kind of story, causal selection is a matter of picking out the most *salient* cause in the context of the specific conversation or

inquiry at hand, and/or ignoring those causes that are taken for granted by all parties and hence do not need to be mentioned. We'll call this strategy *Pragmatics.*

The chapter proceeds as follows. In §2 we give a general overview of the basic issue surrounding egalitarianism, drawing particular attention to its connection with the metaphysics of causation and specifically with the claim that causation is—as we'll put it—a 'natural' relation. In §3 we survey a range of psychological mechanisms that have been proposed in the x-phi literature for explaining causal selection and discuss how they interact with *Defective Mechanism*. Our basic point will be that, while many authors explicitly take their proposed mechanism to be either defective or not, in most cases they appeal to broadly philosophical rationales for endorsing or rejecting egalitarianism. So even if the question of what the relevant mechanism *is* is empirically tractable, it does not follow that the egalitarianism issue is itself empirically tractable. In §4 we turn to *Pragmatics*. We argue that the methodologies that tend to be employed in causal selection studies—specifically, the particular questions that are asked and the ways in which answers are elicited—are not generally well suited to gaining any traction with respect to *Pragmatics*, and we make some suggestions concerning how the debate might be rendered more empirically tractable. In §5, we very briefly sum up.

Two brief comments on what follows are in order. First, our perspective throughout is one that is broadly sympathetic to experimental philosophy; in particular, the fact that we find the existing x-phi literature on causal selection to be less decisive than one might hope or expect when it comes to egalitarianism is not intended as a criticism. After all, very many of those interested in causal selection are not—we take it—especially interested in egalitarianism. But we *are* interested in egalitarianism, and we are therefore interested in how experimental philosophy might be deployed to shed some additional light on the issue.

Second, the topic under discussion will be 'token'—or, as is now more commonly called, 'actual'—causation: that feature of the world that makes true claims of the form '*x* caused *y*' (or alternatively—and importantly, as we'll later see—'*x* was *a* cause of *y*', or '*x* was *the* cause of *y*') true, where *x* and *y* are particular events, facts, states of affairs or whatever. Actual causation contrasts with 'type-level' or 'structural' causation. Thus 'Jane's throwing the brick at the window caused it to smash' is a claim about actual causation, while 'throwing bricks at windows causes them to smash' is a claim about type-level causation.

2. Causal egalitarianism, conceptual analysis, and metaphysics

We'll start this section with three classic cases that have been used—in either metaphysics or experimental philosophy or both—in the context of discussions relating to causal selection and egalitarianism. First, one of the cases discussed in metaphysics involving causation by absence:

> *Dead Plant:* Before Jon went on holiday, his neighbour Jack promised to water his houseplant on Wednesday. Come Wednesday, Jack completely forgot about the plant. He came home from work and planted himself in front of the TV as usual. As it happens, the Queen—who has no idea that Jon or his houseplant even exist—spent the day watching TV too. Jon's plant died. Had Jack watered it, as promised, it would have survived. But also, equally: had the Queen watered it, it would have survived.[3]

Intuitively—so it is often claimed—Jack's failure to water the plant was a cause of its death, but the Queen's failure was not.

Another case from metaphysics concerns the effect that context allegedly has on causal selection:

> *Venusians:* A lightning strike has occurred in a remote area on Earth, resulting in a forest fire. The forest fire counterfactually depends on (at least two) things: the lightning strike, and the presence of oxygen in the forest. Some (Earthling) forest rangers are trying to get to the bottom of what caused the fire. As it happens, some Venusian scientists are observing the event through their high-powered telescope, and they are also trying to get to the bottom of what caused the fire. Lightning strikes (we are supposing) are a very common occurrence on Venus; the presence of oxygen, not so much.[4]

It's often claimed that the forest rangers would want to say that the lightning caused the fire while the presence of oxygen didn't—it is instead a mere background condition that enabled the fire to occur—despite the fact that the fire counterfactually depends on the presence of oxygen just as much as it does on the lightning strike. It's also claimed (though this is sadly

[3] See for example Schaffer 2000: 295; Beebee 2004: 295; Menzies 2004: 145; McGrath 2005: 126; Sartorio 2010: 262-3.
[4] See for example Putnam 1982: 150; Schaffer 2013: 37-8 & 52; Montminy and Russo 2016: 65-6 & 72; Bebb 2022: 28-9.

untested due to the lack of suitable experimental subjects) that the Venusians would want to say the presence of oxygen, and not the lightning, caused the fire.

Third, much of the work done on causal selection in x-phi centres on the discussion of the now-famous pen vignette:

> *Pen*: 'The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take the pens, but faculty members are supposed to buy their own. The administrative assistants typically do take the pens. Unfortunately, so do the faculty members. The receptionist has repeatedly emailed them reminders that only administrative assistants are allowed to take the pens. On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later that day, the receptionist needs to take an important message … but she has a problem. There are no pens left on her desk.' (Knobe and Fraser 2008: 443)

Knobe and Fraser presented this vignette to a group of 18 undergraduate students. They were then asked to what extent they agreed/disagreed with the claims that Professor Smith caused the problem, and that the administrator caused the problem. On average, the subjects tended to agree with the former and disagree with the latter. Knobe and Fraser conclude from this that moral judgements have an effect on which candidate causes we select: we tend to select events whose occurrence violates some moral or social norm.

This phenomenon has since become known as the Knobe Effect, and many subsequent studies have shown it to be highly replicable amongst a much broader range of subjects (Hitchcock & Knobe 2009; Alicke, Rose and Bloom 2011; Phillips et al. 2015; Clarke et al. 2015; Samland & Waldmann 2015, 2016; Samland et al. 2016; Kominsky et al. 2015; Kirfel & Lagnado 2017; Livengood et al. 2017; Icard et al. 2017; Henne et al. 2017; Sytsma 2019; Gerstenberg & Icard 2020; Güver and Kneer 2023). Several of these studies reuse *Pen*, but many also use their own vignettes—and such studies have consistently confirmed the Knobe Effect. It is no exaggeration to say that this is one of the most robust findings from all of x-phi. In addition, many of the studies (see also Hilton and Slugoski 1986) suggest that selection also occurs in cases where the norm that is violated is not a social, moral or legal norm but a statistical norm—that is, people tend to select a statistically abnormal cause over a statistically normal one.

The idea that we 'select' norm-violating causes can be applied not just to *Pen* but also to *Dead Plant* and *Venusians.* The Queen's omission was *normal*, both in that it was not statistically irregular—she never showed even the slightest inclination to nip up to Manchester and water Jon's houseplant—and in that (unlike the omission of Jack, who broke his promise) it also violated no moral norm. In *Venusians*, what distinguishes the forest rangers from the Venusians—and hence what distinguishes their (alleged) difference when it comes to the causal claims they assert and deny—is that oxygen is normal on Earth and abnormal on Venus; and vice versa in the case of lightning. While there may be other drivers for causal selection besides norm-violation—spatial or temporal proximity, for example—our concern in the rest of the paper is with norm-violation.

As we said in §1, our primary focus is on what we're calling 'egalitarianism':

(EC) Our ordinary concept of causation is egalitarian (that is, non-selective).

Since what we're ultimately interested in is metaphysics, we're interested in (EC) because of its relationship to the following metaphysical thesis:

(NR) Causation is a 'natural' relation—that is, a relation whose obtaining is independent of our interests, inquiries, conversational contexts, and so on.

In the rest of this section, we say a bit more about what these theses amount to, how they fit together, and how they interact with the x-phi literature on causal selection.

We'll start with (EC). Whether or not the phenomenon of causal selection undermines (EC) depends on what, exactly, 'selection' amounts to. Suppose that people's (including Venusians') inclinations to assert and refrain from asserting—or even, perhaps, denying—the relevant causal claims correspond to their *beliefs* or *judgements* about which events or actions are and are not causes of the effects in question. Suppose, moreover, that the relevant judgements are *true*: Jack's omission was a cause of the plant's death while the Queen's was not, and so on. In that case, (EC) is false: according to the competent deployment of our ordinary concept of causation, Jack's omission and the Queen's are not, causally speaking, on a par.

There is undoubtedly something to be said for an inegalitarian conception of causation. Hitchcock (2017), for example—in the context of developing an account of 'actual causation' that accounts for causal selection within the causal modelling tradition (see also

Hitchcock and Knobe 2009)—says that 'a central function of identifying actual causes is to identify interventions that will work in combination with other interventions that are feasible, expected, and desirable. These interventions are the kinds that serve to make a system more normal, in the relevant senses' (Hitchcock 2017: 129). So, for example, in a case where causal selection selects the moral norm violator (e.g. Professor Smith in *Pen*) and not the non-violator (the administrator), it's Professor Smith's behaviour that constitutes the locus of desirable intervention: our aim being a harmonious working environment where nobody has a pen crisis, that aim is best achieved by everyone sticking to the rules. One might take this philosophical rationale—alongside the psychological reality of causal selection—to be a good reason to deny (EC).

On the other hand, there is also something to be said for an *egalitarian* concept of causation. Specifically, it cleanly separates what's going on out there in the world from the various kinds of normative considerations that drive causal selection, and one might reasonably hold that such a separation is useful for a variety of purposes. For example, an egalitarian conception would seem to be required for certain kinds of consequentialist views in ethics: if whether or not agent *S* did the right thing in *A*-ing depends on the causal consequences of that action, then there needs to be some evaluatively neutral way of determining what those consequences were. Or in a legal context it might be important to figure out what the consequences of the defendant's negligence were—that is, what harms they caused—prior to assessing whether or not they are legally culpable for it (Stapleton 2008). Or one might think that the egalitarian concept is appropriate in at least some scientific enquiries. According to (EC), that egalitarian concept *is* our ordinary concept of causation. So whatever selection is doing, it is *not* distinguishing between events (or absences, or states of affairs, or whatever) that—according to our ordinary concept—are and are not genuine causes of the effect in question.

What about (NR)? We are not here using the term 'natural' here in the Lewisian sense that is intended to pick out an elite class of properties and relations—perhaps those that make for genuine similarity, or that feature in the formulation of our best scientific theories (Lewis 1983). Rather, we are using it in something like Peter Strawson's sense: 'causality is a natural relation which holds in the natural world between particular events or circumstances, just as the relation of temporal succession does or that of spatial proximity' (1992: 109; see also Bebb 2022). So we're understanding (NR) to amount to the claim that causation is a feature

of the natural world: a feature that obtains independently of our interests and inquiries and therefore independently of the broadly normative considerations that drive causal selection. [5]

Take *Pen*, for example. If Professor Smith's taking the pen, and not the administrator's, caused the problem, then the causal facts at issue depend on local norms regarding the permissibility of pen-taking. Those norms depend on human interests, broadly conceived: the fact that administrators but not professors are permitted to help themselves to pens is not a feature of the physical goings-on in the vicinity of the reception area during the time in question. Had the norm been that professors but not administrators are allowed to take pens, more or less the same sequence of events would presumably have induced people to think that it was the administrator's taking the pen, and not Professor Smith's, that caused the problem. Had the Queen, and not Jack, promised to water his plants, people would doubtless be inclined to assert that her failure to do so, and not Jack's, caused their death—even though the facts on the ground might have transpired in just the way that they actually did, with both the Queen and Jack going about their business without giving the plants a second thought. And so on. So we take (NR) to be incompatible with the claim that social or moral norms make a difference to what causes what.

As we said above, there is evidence that selection (at least sometimes) depends not on the violation of social, moral or legal norms but on *abnormality,* which one might take to be a mind-independent matter. However, while this may be so for some ways of thinking about abnormality—in a biological sense that is defined in terms of proper function, for example (Wachbroit, 1994)—some kinds of abnormality that appear to drive causal selection are arguably inquiry-dependent. In particular statistical abnormality generally depends on choice of reference class, as per *Venusians*. So we take (NR) to be incompatible with the claim that facts about what is or is not statistically normal make a difference to what causes what as well.

According to (NR), then, causation is 'out there in the world', ready for us to uncover through empirical inquiry. (NR) is thus a metaphysical thesis: a thesis about the kind of thing causation itself is. And (NR) goes hand-in-hand with an egalitarian conception of causation, since the claim that causation is a natural relation can only be true if the concept of causation

[5] Our interests and inquiries might of course make a difference to the causal order in the sense that Helen's interest in securing a cold beer might manifest in a desire for a cold beer, which in turn might play a causal role in her ordering one; or Jon's inquiry into the death of his plant might manifest in various actions that have effects. That isn't the sense at issue here.

deployed in that claim is egalitarian.[6] Hence, according to (NR), causal selection does not identify the genuine causes of an event; rather, it selects *some* genuine causes and deselects others.

Whether or not they have explicitly or implicitly taken our *ordinary* concept of causation to be the one they are using, traditionally most metaphysical accounts of the nature of causation endorse (NR)—and therefore (at least implicitly) take the concept they *are* using to be egalitarian. Consider Lewis's classic counterfactual analyses of causation. Whether an effect *e* counterfactually depends on (or is linked by a chain of counterfactual dependence to) *c*, and hence is a cause of *e*, is insensitive to whether or not *c* violated a norm or is relevant to our current inquiry or whatever: the relevant facts about counterfactual dependence turn only on facts about regions of spatiotemporal match and the laws in the actual and the closest ~*c*-worlds. Accordingly, Lewis says:

> We sometimes single out one among all the causes of some event and call it 'the' cause, as if there were no others. Or we single out a few as the 'causes,' calling the rest mere 'causal factors' or 'causal conditions.' Or we speak of the 'decisive' or 'real' or 'principal' cause. We may select the abnormal or extraordinary causes, or those under human control, or those we deem good or bad, or just those we want to talk about. I have nothing to say about these principles of invidious discrimination. I am concerned with the prior question of what it is to be one of the causes (unselectively speaking). My analysis is meant to capture a broad and nondiscriminatory concept of causation. (1973: 558-9)

Here it is clear that Lewis is describing the phenomenon of causal selection, and—minimally—that, whatever the *ordinary* concept of causation may be, *his* interest is in the egalitarian concept of causation. (We think Lewis does endorse (EC), but we'll park this for now.) And of course, as just explained, this goes hand-in-glove with the fact that Lewis's counterfactual account of causation is one according to which causation is a natural relation. Similarly, we take it that other traditional accounts of causation—such as accounts according to which causation is the transfer of some conserved quantity such as energy-momentum

[6] A caveat: if causation is a *scalar* relation—if *x* can be more or less of a cause of *y*—and if the determinants of the extent or degree to which *x* is a cause of *y* are themselves 'natural' in our sense, then there is a sense in which (NR) is true and yet not all causes are on a par. We return briefly to this issue in §4.

(Dowe 2000, 2004) or the manifestation of a power (Mumford and Anjum 2010)—are deploying an egalitarian conception of causation and, correspondingly, endorse (NR).

Causal selection thus threatens to put indirect pressure on (NR): *if* (a) causal selection is incompatible with (EC), and *if* (b) the feature of the world that is the topic of investigation in the metaphysics of causation is supposed to be the very same feature that is picked out by our ordinary concept, then (NR) is false.[7]

One way to resist that pressure is of course to deny (b), and to claim that there is *a* concept of causation—one that bears a close enough relationship to our ordinary concept to merit the name—that is egalitarian; it's just that that's not our ordinary concept of causation. Hitchcock and Knobe, for example, claim that 'egalitarianism is entirely appropriate at the level of what we call causal structure' (2009: 593; see also Hitchcock 2007). One might even hold—in the spirit of conceptual engineering—that the egalitarian concept is a *better* concept of causation than the inegalitarian concept we (allegedly) actually deploy in everyday life, precisely because it makes a sharp—and useful, for one or more of the reasons surveyed earlier—distinction between what's out there in the world and what's normal or contextually salient or whatever. One might even go so far as to claim that the ordinary concept is simply irredeemably *defective*—'surviving, like the monarchy, because it is erroneously thought to do no harm', as Bertrand Russell put it (1912–13, 1).[8] Alternatively, one might hold that the egalitarian concept is perfectly good in its place—when we're doing metaphysics, say—but the ordinary concept is better suited to the kinds of inquiry we are normally interested in in our everyday lives.

We shall ignore these kinds of proposal in what follows, however, since we take it that their proponents are simply not especially interested in the ordinary concept of causation—and, for the purposes of this paper, we *are* interested in it. Our main concern in the rest of the paper, then, will be (EC). To focus attention, let's consider some examples of views in the metaphysics of causation that endorse and reject (EC).

Starting with the former, we take Lewis's view of causation to constitute endorsement of (EC). As we saw above, Lewis says: 'My analysis is meant to capture a broad and nondiscriminatory concept of causation'. On its own, this is something that someone who

[7] But not *only* if, since one might have other reasons for denying (NR)—for example if one took causation to involve a projection of the mind (Ramsey 1929; Price 2007; Beebee 2015).

[8] Russell himself thought we should do away with causal talk all together—but other less radical options are of course available.

rejects (EC) could happily say—in particular someone who thinks that, by virtue of indulging in 'invidious discrimination', the ordinary concept of causation is defective and should be replaced by one that is 'nondiscriminatory', that is, egalitarian. We take it that this is not Lewis's view, however. His project when it comes to the metaphysics of causation is one that starts with conceptual analysis, and he takes conceptual analysis of a given phenomenon *X*—in general—to be a matter of uncovering the ordinary concept of *X*. And he moves from conceptual analysis to metaphysics by simply taking it for granted that—unless we have *very* good reasons to think otherwise—our ordinary 'opinions' are, by and large, correct. That is to say, having figured out that the ordinary concept of causation is that of (transitive chains of) counterfactual dependence, and hence is egalitarian, that's what causation *is*.

As an example of a view that denies (EC), consider contextualism, as defined and defended by Peter Menzies:

> Take any objective substructure of events and relations, whatever it may be, this pattern cannot determine the truth-conditions of a causal judgement, because its truth-value can vary from one context to another, depending on how a certain contextual parameter is set. The very same pattern of objective relations, viewed from within one context, may support a causal judgement, but, viewed in another context, may fail to support the judgement. (Menzies 2007: 192-3)

Contextualism is clearly incompatible with (EC): if the truth of a causal judgement can vary with context, then the truth of that causal judgement is determined, in part, by our own interests and inquiries (our choice of 'contextual parameters') and not with the objective situation under investigation.

One way in which broadly contextualist views have been developed in recent years is via causal models. A causal model represents various features of the actual situation, and counterfactual variants of those features, as values of variables. The model then generates causal truths according to whether toggling between the values of a given variable (the putative cause) changes the value of another (the effect). Crudely: if this variable had taken one of the other values represented in the model, would the effect variable also have taken a different value?

Roughly speaking, there are three features of causal models that can be exploited to make them well suited for enshrining causal selection. One is that they are often taken to be subject to an 'aptness' constraint according to which 'variables should not be allotted values

that we are not willing to take seriously' (Blanchard & Schaffer 2017: 182). One might plausibly take the possibility of the Queen's watering Jon's plant to fall foul of the aptness constraint, in which case there will be no apt model for *Dead Plant* where the Queen's watering the plant or not (Z) is permitted to take the value Z=1 (she waters it).

A second feature is the fact that we can construct different causal models of the same target system by including or excluding various (serious) values of the variables. A 20mg-or-more dose of poison is fatal; Assassin delivers a 25-mg dose. Did Assassin's action cause Victim's death? Well, if we deploy a model that only includes (say) 20mg, 25mg and 30mg as values of the dosage variable X, the answer is 'no': Victim would have died whichever of those values X had taken. By contrast, if we deploy a model that includes (say) only 25mg and 0mg, the answer is 'yes'.

Finally, a third feature of causal models is that we can (though it's controversial whether we ever *should*) fix the values of variables at *non-actual* values. This is one way to deal with tricky pre-emption and overdetermination cases: Billy and Suzy both throw rocks at a window, but Suzy's gets there just before Billy's and hence Suzy's throw causes the window to smash while Billy's does not. If we hold fixed Billy's throwing-or-not, BT, at BT=1 (Billy throws), we get no counterfactual dependence between Suzy's throw and the smashing; however, if we hold fixed Billy's throw at the non-actual value BT=0 (no throw), we get the right result.[9] This basic strategy has also—more controversially—been used to deal with a range of other difficult cases, including *Plant Death* (Menzies 2004, 2006, 2007; see also Blanchard and Schaffer 2017). The idea is that we hold fixed off-path variables at their 'most normal' values—whether or not those values are the actual values. This approach has come to be known as 'default-relativity'.

Each of these features of causal models (if deployed) constitutes an implicit rejection of (EC). While the aptness constraint (the first feature above) is not strictly contextualist—since, we take it, aptness is a universal constraint on causal models and is not supposed to vary from one context to another—it is nonetheless a constraint that makes the truth or falsity of a causal claim depend on what we are willing to take seriously; and that is of course a feature of us, and not a feature of (for example) the situation *viz a viz* the demise of Jon's plants. The second feature clearly (we think) delivers a contextualist position: whether it

[9] The same is true, of course, of Suzy's throw—so obviously we need a way to ban holding the Suzy variable ST fixed at ST=0. Without going into the detail, the basic idea is to distinguish between 'off-path' and 'on-path' variables and to ban the holding-fixed of on-path variables. In this case, BT is off-path and ST is on-path.

comes out true or false that Assassin's delivering the 25mg of poison caused Victim's death depends on our choice of model, and that choice is not forced on us by the nature of the situation being modelled. Thus, for example, Assassin might be kicking herself for wasting precious poison. '*That* dosage didn't cause Victim's death', she might think; 'he would have died just the same if I'd given him a bit less'. A court of law would of course invoke a different model.

Finally, the third feature—default-relativity—also constitutes a rejection of (EC), since what is 'normal' is (at least very often) either determined by social, moral or legal norms or, in the statistical case, determined by the selection of a reference class. (Nine times out of ten, Jane eats her lunch at her desk. Today, however, she's accepted a lunch invitation and is eating at the café. Is her absence from her desk at lunchtime normal? Well, it only happens on 10% of weekdays, so no. On the other hand, it happens 100% of the time she's accepted a lunch invitation, so yes.)

A final view we'll briefly consider is the contrastivist position, according to which causation is a ternary or quaternary relation rather than a binary one (see Hitchcock 1996; Maslen 2004; Schaffer 2005, 2013; Northcott 2008). On this view, the sentence '*c* caused *e*' expresses an incomplete proposition, where the complete proposition is (in the quaternary version) <*c* rather than *c** caused e rather than *e**>, and *c** and *e** are counterfactual alternatives to *c* and *e* respectively. Which counterfactual alternative is in play is determined by conversational context. Thus, for example, while 'delivering 25mg of poison caused Victim to die' is true in the context of the murder trial but (let's assume) false in the context of the Assassin berating herself for wasting her precious poison, this is not because the facts about which causal *relations* obtain vary with context; it's rather because the different contexts select different counterfactual alternatives—no poison at all in the first case and a bit less poison in the second—and hence different causal relations. It is context-independently true that using 25mg rather than no poison caused Victim to die rather than survive and, equally, context-independently false that using 25mg rather than 20mg caused Victim to die rather than survive. Those two propositions simply pick out different causal relations, both of which were objectively present in the situation under discussion.

Contrastivism is generally conceived as a version of contextualism; we think it's better not to, however, on the grounds that contrastivism is, but contextualism is not, compatible with (NR): the contrastivist, unlike the contextualist, is free to hold that causation itself—the quaternary relation—is entirely independent of norms, etc. (see Bebb 2022). It's a little unclear where this leaves (EC). On the one hand, insofar as our concept of causation

refers to a selection-independent four-place relation, it's non-selective; on the other, our causal *claims*—two-place claims of the form '*c* caused *e*'—*are* selective. Indeed, Jonathan Schaffer (2005) explicitly motivates contrastivism via (*inter alia*) its ability to solve what he calls the 'paradox of selection', and suggests that '[p]erhaps the idea of a "broad and nondiscriminatory concept [of causation]" is a *philosophers myth*' (2005: 344). It is noteworthy, however, that contrastivism doesn't actually line up very well with the x-phi evidence on causal selection. It cannot, for example, distinguish between Professor Smith's and the administrator's actions in *Pen*, since (presumably) the only possible contrast in both cases is that they *didn't* take the pen; similarly for *Plant Death*. So even if we take contrastivism to constitute a rejection of (EC), the contrastivist will still need to explain *some* cases of causal selection by adopting one or more of the strategies we come to in the next two sections.

That completes our survey of various kinds of view that accept and reject (EC). It's time to get back to the psychology of causal selection and see what light, if any, x-phi work in this area can cast on the disagreements between those views.

3. Causal selection and psychological mechanisms

As we've said, causal selection puts pressure on (EC)—the claim that the ordinary concept of causation is egalitarian—and therefore, less directly, on (NR), the claim that causation is a natural relation. As we said in §1, there are two broad lines of defence available to the defender of (EC). One might defend the view that the psychological mechanism(s) responsible for causal selection is *defective* in the sense that, in counting non-norm-violators as non-causes, it is prone to delivering false judgements about what is a cause of what. That's the *Defective Mechanism* strategy, which we take to be broadly the same as what David Rose (2017) calls a 'debunking explanation'. Or one might argue that people's inegalitarian responses to the various thought experiments are to be explained not by psychological mechanisms that issue in *judgements* concerning the truth or falsity of the relevant causal claims, but rather by pragmatic phenomena such as conversational implicature or pragmatic presupposition—so those inegalitarian responses are, in effect, sensitive to what is and is not contextually appropriate or salient rather than to what people take to be the causal facts. This is the *Pragmatics* strategy.

When considering the import of the now very large experimental philosophy literature on causal selection for the debate about (EC)—and therefore, indirectly, (NR)—two immediate problems for assessing these two defensive strategies present themselves. The

first, which relates to *Defective Mechanism*, is that much of that literature is explicitly concerned with the project of figuring out which psychological mechanism or mechanisms best explain the experimental data, and this project is often (though not always) somewhat orthogonal to the question of whether (EC) is true. Given the goal of identifying the psychological mechanism(s) at work, one might not be especially interested in the question of whether that mechanism is defective—in whether it is prone to generating *mistaken* judgements—or whether instead, *contra* (EC), the mechanism manifests *competence* with respect to the concept of causation. And it is not always easy to see how we might get an independent grip on whether or not that mechanism is defective—independent, that is, of simply taking a stand on whether or not (EC) is true, which of course in the current context is precisely what is under dispute. As Rose says with respect to his own preferred account of causal selection: 'Where I see a glaringly defective epistemic process, others see an epistemically appropriate process' (2017: 1327).

The other problem, relating especially to *Pragmatics*, is that the methods used—in particular both the questions that experimental subjects are asked and the ways in which they are asked—are very often not designed in such a way as to adjudicate whether or not pragmatic factors are at work in driving subjects' responses. This is, presumably, a general feature of x-phi methodology, but it is a particular problem in the case of causation given that a pragmatic interpretation of those responses is very much a live option when it comes to the philosophical debate surrounding the ordinary concept, and the metaphysics, of causation. In the rest of this section we discuss some examples from the x-phi literature that highlight the first of these problems; in §4 we discuss the second.

In the rest of this section, then, we'll survey some of the psychological mechanisms that have been proposed for explaining causal selection and consider how they bear on the *Defective Mechanisms* strategy for defending (EC). One kind of psychological explanation, which we'll follow Alicke *et al.* (2011) in calling the 'evaluative perspective', explains causal selection by appealing to the normative evaluation of the act or event whose causal status is at issue.[10] Thus, for example, Alicke and colleagues explain the fact that (for example) Professor Smith gets a higher causal rating than the administrator by appealing to a desire to blame. Very roughly, in cases where we have two candidate causes of some outcome, one of which involves an agent we want to blame for their action, our natural and automatic

[10] Henne *et al.* (2021) call accounts of this general kind—focussing on blame and praise—as 'social-cognition explanations'.

retributive instincts cause us to upgrade that agent's control over the situation, and hence their causal status, and downgrade the causal status of the other candidate cause (see Alicke, 1992, 2000; Alicke and Rose 2012; Alicke, Rose and Bloom 2011; Rose 2017). According to the evaluative perspective, norm violation is relevant to causal selection in that at least part of what drives judgements of blameworthiness is the violation of moral norms.

Is the evaluative perspective congenial to the *Defective Mechanisms* strategy? Alicke, Rose and Bloom (2011) are somewhat equivocal on that question. On the one hand, they say that 'evaluative reactions generally provide the opportunity for pervasive biases in causal judgement' (2011: 691). On the other, they claim that the evaluative influences on causal judgement 'are essential components of lay behavioral analyses, because they stem from observers' most fundamental motives of discerning which objects, events, and people are likely to facilitate their goals and well-being, and which endanger their prospects' (2011: 693)—and it's unclear why (unless we presuppose that (EC) is true) we should not think that our causal judgements are *appropriately*, as opposed to defectively, attuned to those motives. Rose (2017) is decidedly unequivocal, however: 'the role of moral considerations in causal judgements … is epistemically defective' (2017: 1327). However, his grounds for this claim largely consist in the kinds of reasons for favouring an egalitarian conception of causation that we briefly canvassed in §2 above.

One limitation of the evaluative perspective is that it does not explain causal selection in non-agential cases, where the desire to blame is not (at least *prima facie*) a very promising starting point. Rose fills that lacuna by positing a second, distinct mechanism, which he calls the 'agentive process'. The idea, based on evidence that people tend to engage in 'promiscuous teleology', is that we have a tendency to—as Paul Bloom nicely puts it—'chew over the natural world with our social mode of understanding' (2007: 150) and attribute agency to natural objects: rocks, plants, and so on. Rose ran an experiment that suggests that 'non-obvious agentive considerations—as measured by Gaia beliefs—directly affect causal judgements for non-living things' (2017: 1335), where Gaia beliefs are beliefs such as 'the Earth is alive', 'Nature is driven to preserve living things', etc. (2017: 1334). Rose—very plausibly—takes the agentive process to be epistemically defective when deployed in situations where 'the folk are mistaken to view the events under consideration teleologically [that is, in agentive terms]' (2017: 1348).[11] However, he is neutral on whether causal

[11] In fact Livengood, Sytsma and Rose—in the context of defending the responsibility view, to be discussed below—take the tendency 'to take an agentive perspective on nature as a whole … to provide reason to expect

selection on broadly teleological or agentive grounds is epistemically defective in cases where the folk are *not* mistaken about teleology.

In both cases, then—the evaluative perspective and (at least in cases that do not arise from *promiscuous* teleology and are hence not *obviously* defective) the agentive-process explanation—it is unclear whether the psychological mechanism in question should be thought of as defective. What about other proposed mechanisms? Well, let's briefly consider a pair of proposals: the *responsibility view* and the *accountability hypothesis.* According to the responsibility view (Livengood and Sytsma 2020; Sytsma 2019, 2020, 2022; Sytsma and Livengood 2021; Sytsma *et al*. 2012; Livengood *et al*. 2017), the ordinary concept of causation is the concept of what—to distinguish it from the egalitarian concept presupposed in much of the metaphysics literature—Sytsma and co-authors call *responsibility.* As Sytsma and Livengood explain it: 'the default concept of causation at play in ordinary causal attributions has some indispensable evaluative content alongside whatever nonevaluative content it has … this concept is a kind of thick ethical concept akin to our concept of responsibility' (2020: 48). Livengood and Sytsma explicitly thus take causal selection to be the result of a non-defective psychological mechanism.

By contrast, according to a proposal of Samland and Waldmann (2015, 2016)—the accountability hypothesis—ordinary causal attribution is sensitive to two 'layers of description of an action': 'The *causality layer* lies at the bottom and refers to the causal mechanisms connecting actions and outcomes in the scenarios. For example, in the pen vignette both agents initiate behaviors that lead to the removal of pens. The causality layer is a subset of the morally charged second layer, the *accountability layer*' (2015: 2093). And, they say, 'a possible alternative explanation of the findings suggesting that causal inferences are influenced by norms might just be that subjects interpreted the test questions as a request to assess moral accountability' (2015: 2093).

This proposal is clearly pretty closely related to the responsibility view, in that both understand our causal judgements about cases involving norm-violating agents as implicitly responding to the question of who is responsible or accountable for a given outcome. The

---

that people will hold non-agents responsible' (2017: 292, n.40). According to the responsibility view, such attributions of responsibility underpin causal attributions and there is nothing defective about that. So even the claim that there is something defective about causal attributions if they depend (in non-agential cases) on taking 'an agentive perspective' is disputed.

difference between the two is that Livengood and Sytsma view responsibility as baked into the concept of causation, and so the relevant judgements are straightforward (non-defective) *causal* judgements that arise from a competent application of the concept. By contrast, Samland and Waldmann maintain a firm divide between accountability and causation. Despite being couched in the language of causation, the relevant judgements are judgements about accountability, for which causation is a necessary but not a sufficient condition. We take it that Samland and Waldmann do not think there is anything *defective* about the mechanism delivering the judgements—they are, after all, correct attributions of *accountability*; nonetheless, unlike the responsibility view, the accountability hypothesis is broadly consistent with (EC).[12]

We shall not attempt to adjudicate this dispute here. Our main point in describing it is to highlight the fact that identifying a plausible psychological mechanism that delivers causal selection is one thing; making a call on whether that mechanism issues in the competent deployment of the ordinary concept of causation or is instead defective is quite another.

Finally, we'll consider what Henne *et al.* (2021) call 'modal explanations', and in particular Hitchcock and Knobe's (2009) proposal, which takes norm violation—in the most general sense, including not just moral, social and legal norms but also statistical norms and norms of proper functioning—to directly drive causal selection. Their account provides a *modal* explanation of causal selection because norms determine which counterfactual situations we take to be relevant to our causal attributions: roughly, the worlds where what happens is what *normally* happens are more relevant than those where what happens *isn't* what normally happens. In *Pen,* for example, the administrator's taking a pen is *normal,* and 'if some normal event occurs, we may never get around to considering the counterfactual situation in which some more abnormal alternative occurs instead. Even if this situation is one in which the outcome is different, the normal event will not be judged a cause of the outcome' (2009: 601). The idea here, then, is that while it is *true* that the problem would not have happened had the administrator not taken a pen, that fact is irrelevant to the attribution of causation because the counterfactual scenario it involves is not a normal one. By contrast, Professor Smith's taking the pen is abnormal—in the sense that it violates the office rule—and

[12] There is a sense in which the accountability hypothesis belongs with the *Pragmatics* strategy discussed in the next section, in that it takes responses in x-phi studies *not* to straightforwardly express subjects' beliefs about what caused what. We briefly return to this in §4.

so the counterfactual situation in which he doesn't take the pen *is* normal, and hence is considered relevant.

Hitchcock and Knobe are clear that they do not take the psychological mechanism to be defective. Indeed, they offer a robust philosophical defence of the legitimacy of taking 'normal' counterfactual scenarios, and not abnormal ones, to be relevant in determining causal judgements: doing so grounds our judgements about which interventions would be most effective—namely, interventions on abnormal events (see their 2009, §10; see also Hitchcock 2017 and §2 above). The account therefore dovetails nicely with the 'default-relativity' approach in the metaphysics of causation (see §2).

On the other hand, consider Blanchard and Schaffer's '*Heuristics-and-Biases Strategy*' (2017: 208-11). They propose exactly the same mechanism as Hitchcock and Knobe, but take it to be a *bias*. Like Hitchcock and Knobe, they appeal to Tversky and Kahneman's work on cognitive heuristics (1973), as well as later work by Kahneman and Miller (1986); but they draw attention to the fact that the 'availability heuristic'—in this context, the availability of counterfactual scenarios—is a source of bias in other domains in which it operates (e.g. in judgements about probabilities and frequencies). This being so, someone who takes availability to be a bias rather than (as Hitchcock and Knobe do) a manifestation of competence with respect to the concept of causation has '*a ready explanation for the generality and systematicity of normative influences*. After all, she is positing a single background feature of cognitive performance playing a role in all of these domains' (Blanchard and Schaffer 2017: 211). The dispute between Hitchcock and Knobe on the one hand and Blanchard and Schaffer on the other makes vivid the gulf between positing a psychological mechanism that explains causal selection—on which, in this case, both parties to the dispute agree—and making a call on whether or not that mechanism is defective, on which they very much do not agree.

As the above brief and incomplete survey indicates, mounting a successful defence of (EC) —the claim that the ordinary concept of causation is egalitarian—via the *Defective Mechanism* strategy is certainly a live option as things currently stand. Of course, the debate about the psychological mechanism(s) that best explain causal selection is ongoing, and there is a burgeoning literature that attempts to experimentally adjudicate between the various proposals; and we have no idea how the debate will pan out from an empirical point of view. But we think there are good inductive grounds for thinking that, however it pans out, the

viability of *Defective Mechanism* will—as at present—continue to be significantly empirically underdetermined.

That said, we do want to highlight a particular issue that has not garnered as much attention as it might. Danks, Rose and Machery (2014) and Samland and Waldmann (2015) present evidence that causal judgements are only influenced by norms in cases where the causal structure of the situation is learned by description rather than by experience. In Samland and Waldmann's study (2015, Experiment 1), all participants were given a cover story involving Tom the gardener, who has two insecticides available (call them *X* and *Y*), but has read that using both together can kill the plants. He therefore forbids his two assistants—Alex and Benni—from using *Y*. Benni, however, continues use *Y*. One day, Tom discovers that some of the plants have dried up. In the description condition, participants were then told that the plants were in a bed sprayed by both Alex and Benni—and that, while Alex had used the allowed chemical (*X*), Benni had used the forbidden one (*Y*). In the learning condition, participants were told that Tom would like to 'investigate the relationship between the use of the chemicals and the shriveled plants' (2015: 2094), and were then shown a series of slides (ten of each: no chemicals and no plant; *X* only and plant healthy; *Y* only and plant healthy; *X*&*Y* and plants dried up). In the description condition Benni's spraying was given a higher rating than Alex's; in the learning condition there was no significant difference between the causal ratings.

Samland and Waldmann take these results as evidence for their accountability hypothesis, discussed above: their claim is that the learning condition emphasises the 'causal mechanism' at work, while the description condition cues participants to 'interpret the test question as an attempt to assess accountability' (2015: 2094). Danks *et al.* are merely concerned to challenge what they call the 'Ubiquity Thesis' that 'normative considerations … fundamentally influence causal cognition in general' (2014: 262). They conclude that it 'may well be that the influence of normative considerations is limited to a narrow, linguistically-mediated form of causal cognition' (2014: 263). If the finding that norm-violation only affects causal judgement in cases where the causal structure is described rather than learned is robust, that would constitute grist to the egalitarian's mill. After all, participants must be getting it wrong in *one* of the two conditions whether or not egalitarianism is true, since the two conditions (described and learned) are representing the same situation. If Danks *et al.* are right, then it would seem that the inegalitarian is in no better a position than is the egalitarian when it comes to attributing widespread error to our ordinary judgements; it's just that the use of descriptive vignettes renders this fact invisible.

4. Pragmatics and x-phi methodology

The discussion of §3 largely assumed—as much of the x-phi literature does—that responses in x-phi studies straightforwardly measure the strength of participants' inclinations to *judge* whether or not (or perhaps to what extent) one event or action *x* is a cause of another, *y*. That is, the x-phi literature tends to assume that the responses elicit participants' *beliefs* about the causal status of the candidate causes. Samland and Waldmann's accountability hypothesis is an outlier in this regard, in that it characterises the participants' responses as expressing beliefs not about causation but instead about accountability. Another outlier is Hilton's (1990) account, according to which participants' responses are, in effect, causal *explanations* that are 'constrained by general rules of conversation' (1990, 66)—which Hilton takes to be, roughly, Grice's conversational maxims.

Hilton's account is an instance of the egalitarian strategy we have labelled *Pragmatics*. According to this strategy we should reject the assumption that the participants' responses accurately reflect their causal beliefs, and instead claim that the responses merely measure the strength of participants' *willingness to assert* the statements with which they are presented; and of course one's willingness to assert some statement need not correlate especially closely with what one actually believes, since pragmatic considerations often make us *un*willing to assert things that we nonetheless believe.[13]

In particular, a broadly Gricean story highlights the importance of *salience*—in Grice's terms, the maxim of *Relevance* (Grice 1975: 47). In the context of a given inquiry or conversation, a statement might be believed but irrelevant—and, since participants in the conversation will generally assume that *Relevance* is being obeyed, giving irrelevant information is apt to generate false implicature, as when someone who clearly needs petrol asks you where the nearest petrol station is and you respond, truthfully, that there's one just down the road—while knowing that it is closed. You express a true belief—but, in violating *Relevance*, you generate the implicature that the petrol station is open, since your interlocutor will assume that you have provided information that is relevant to the purpose of the conversation and the information is only relevant if the petrol station is open.

[13] Note that Samland and Waldmann's account does not fit cleanly under the *Pragmatics* heading. On the one hand, they don't take the participants' responses to express *causal* beliefs; on the other, they do take them to express *beliefs* rather than being a mere reflection of their willingness to assert the relevant causal statements.

This Gricean story is, at least *prima facie*, one that can be applied to the vast majority of x-phi studies; and, as we'll see, this renders many of the studies that uncover causal selection inapt for getting any empirical traction on the debate about egalitarianism.

To start, recall Lewis's claim quoted in §2:

> We sometimes single out one among all the causes of some event and call it 'the' cause, as if there were no others. Or we single out a few as the 'causes,' calling the rest mere 'causal factors' or 'causal conditions.' Or we speak of the 'decisive' or 'real' or 'principal' cause. We may select the abnormal or extraordinary causes, or those under human control, or those we deem good or bad, or just those we want to talk about. (1973: 558–9)

Lewis is surely right to claim that we *do* do all of these things, and also right in his implicit claim that this does not entail that we are operating with an inegalitarian concept of causation. After all, nobody (presumably) thinks that any event has just *one* cause—and yet people happily use the locution '*x* was *the* cause of *y*'.

This being so, the wording of the questions that are posed to the participants render some x-phi experiments silent on the issue of whether or not (EC) is true. For example, in one of Alicke's 1992 studies investigating whether 'causal judgements are conflated with ascriptions of blameworthiness' (1992: 369), participants were asked: 'Complete the following sentence. The primary cause of the accident was _____' (*ibid.*). The studies in Alicke *et al.* 2011 asked participants 'how much' one thing caused another, with answers given on a scale from 'very much *the* cause' to 'not at all *the* cause' (our italics).[14] The fact that participants tend to give higher ratings to the act of an agent who is judged blameworthy than for the act of an agent who is not judged blameworthy (as in these cases) does not constitute evidence that the subjects are genuinely deploying an inegalitarian concept of causation, since they could just as well be picking out the *most salient* cause. After all, it would be unsurprising if, in a case where there is an agent who is to blame for some outcome, their own causal contribution to that outcome were more salient (in most contexts) than the contribution of other causes.

Most studies, however, uncover causal selection when asking participants to give Lickert scale ratings for '*x caused y*'—and the egalitarian does not have such an easy time of

[14] See below for a brief discussion of the idea that causation is 'scalar'.

it when it comes to explaining these results, since, unlike '*x* is the primary cause of *y*' or '*x* is the cause of *y*', '*x* caused *y*' does not obviously wear any demand for salience on its sleeve. On the other hand, '*x* caused *y*' might still be read as implying uniqueness—that is, as equivalent to '*x* was *the* cause of *y*'—and so a better test of (EC) would be to use '*x* was *a* cause of *y*' (see below).

Moreover, the egalitarian has some additional moves available in at least some of the studies where the 'caused' (or future-tense 'will cause') locution is used. For example, in the studies reported in Henne *et al.* 2021, participants were instructed: 'Please select the phrase that best completes the following statement about the video you just watched: _____ will cause the ball to reach the goal' (2021: 6; see also p.16). In this kind of forced-choice situation there is no opportunity to signal the judgement that *all* of the choices on the table are, equally, causes, and so again it would be entirely plausible to claim that the participants are selecting the most salient cause.

A related problem is raised by studies such as *Pen*, where half the participants were asked whether the administrator caused the problem and half were asked whether Professor Smith caused the problem, with the latter getting a higher rating than the former. A salience explanation can in principle be given here too: it is not implausible to hypothesise that those participants who were asked the former question were disinclined to give a high rating to the administrator on the grounds that doing so would carry the false implicature that the administrator was blameworthy or violated a norm. If blameworthiness or norm violation is *generally* salient in the context of making assertions about causation where there is a blameworthy or norm-violating agent on the scene then, in a conversational context where there is no opportunity to signal one's view that the administrator has done nothing wrong, asserting that the administrator caused the problem plausibly does carry that implicature. (In fact, Knobe and Fraser (2008) themselves are sensitive to the *Pragmatics* strategy and run an experiment to counter just this kind of move; we come back to this below.)

Similar results to *Pen* have been achieved using non-agential cases where blame and social norms are obviously not in play but where, for example, there are two causal factors, only one of which was 'supposed' to happen (Hitchcock and Knobe 2009: 604-5). Again, participants tended to give a higher rating to the factor that was not 'supposed' to happen. Such cases might seem more problematic for the egalitarian since the idea that blameworthiness or norm-violation by agents is salient by default does not apply. However, again, the egalitarian could argue that cases of norm violation quite generally—including malfunctions and statistical abnormalities—are generally more salient, so that giving a high

causal rating to a factor that was operating entirely as it should (or as it normally does) carries a false conversational implicature unless there is an opportunity to cancel it—and such an opportunity is lacking in *Pen* and other similar studies.

One set of studies whose methodology stands out as unusually sensitive to the debate about egalitarianism comes in Clarke *et al.* 2015. In the first study, not only were participants asked to rate statements of the form '*x* was one of the causes of *y*', but they were divided into three groups. One group was not primed, one was given an explicitly egalitarian prime ('Many things were causes of the collision'), and one was given a more extensive egalitarian prime ('Many things were causes of the collision. And the collision wouldn't have occurred if things had been different in certain ways. For example, it wouldn't have occurred if Greta [the agent who didn't violate a norm by running a red light] hadn't driven into the intersection' (2015: 282)). Clarke *et al.* found that norm violation had a significant influence on people's responses, and that the primes had no effect on this. As they say, these results 'support the view that people are surprisingly resistant, even given extensive suggestion, to agreeing that agents who haven't violated norms might nonetheless be among the causes of some outcome' (2015: 283); on the other hand, they note that 38% of participants 'agreed at least slightly that Greta was a cause of the collision' (*ibid.*).

When it comes to gaining any traction on (EC), the approach of Clarke *et al.* is very much a step in the right direction; however, there is a general concern raised by *any* experimental set-up concerning causal selection where selection is, in effect, enforced—as is the case in the overwhelming majority of studies. In any such case—even when participants are primed, as in the above case—it is open to the egalitarian to claim that, in being asked about the causal status of one and only one prospective cause, participants are disinclined to grant that status to causes that they take to be genuine but not the most salient.

There are various ways in which one might remove enforced selection. In Bauer and Romann's study (2022), for example, participants were asked to give causal ratings for a succession of events along the causal chain leading to someone being shot (the release of the hammer, the hammer striking the cartridge, etc.), where all of the events are listed together on the same screen (2022: 189-90) and all of the participant's ratings are submitted at once. This kind of approach—as well as asking the 'was a cause' rather than the 'caused' question—ought to remove or at least reduce any inclination on the part of participants to single out the most salient cause, since they are not being forced to choose between (as the egalitarian would have it) the causes of the event in question.

An additional proposal is to give participants the explicit option of selecting *all* of the relevant factors as causes. In a recent study by Wesley Buckwalter (m/s), participants were explicitly asked whether the deletion of emails—which required two people to log on to the computer—happened because the norm-violator logged on to the computer, or because the non-norm-violator logged on, or because both logged on; over 90% chose 'both'. Similarly, when asked to rate the statement 'While someone did break the rule, the fact that two people were logged on is what actually caused the computer to delete the emails' on a 7-point agreement scale, the mean response was 6.01.

A final proposal—one that is in fact adopted by Knobe and Fraser (2008) with the explicit aim of testing the *Pragmatics* strategy but has not, so far as we know, been pursued more widely—is to go for a much more direct approach. Using a vignette where both a norm-violator and a non-violator are necessary for the effect, they explicitly asked participants *both* 'what explanation would be most *appropriate* [for a third-party witness to the situation] *in the conversation*', where it would clearly not be salient to mention the norm-violator, and, separately, what that person '*actually believes*' (2008: 445). The non-violator was assigned a high rating for the appropriateness question but a much lower rating than the violator for the belief question, suggesting that norm violation really is affecting judgements about what caused what, rather than merely affecting conversational salience. This result does not bode well for the egalitarian. However the study was very small; it would be interesting to conduct further studies along similar lines.

A final issue that the x-phi literature has not, to our knowledge, addressed so far is what—if any—causal status participants attribute to those events, actions or factors to which they assign low or zero causal ratings. According to *Pragmatics*, participants are not judging those factors to be non-causes but rather *bona fide* causes that are not salient: causes that it would be potentially misleading to mention, or whatever. However the x-phi literature does not, so far as we know, have much to say about what—according to the view that causal selection *does* amount to a judgment about what is and is not a cause—the exact causal status of those low-rated factors is judged to be. One suggestion would be that they are being judged to have *some* kind of broadly causal status—perhaps they are mere 'causally relevant factors' or 'conditions' or nodes in the 'causal structure' or some such. That suggestion is consistent with the idea that we possess *a* causal concept that is egalitarian—the concept of a causally relevant factor, say—where *some* such factors achieve the status of actual cause. But the suggestion seems somewhat unpromising in the light of the fact that we seem, at least on

the face of it, to lack an expression in ordinary English for such a concept ('causally relevant factor', 'condition' and 'causal structure' being, we take it, terms of art).

An alternative suggestion would be that there is no such broadly causal concept and people really do judge that those factors to which they give low or zero causal ratings are not in *any* way causally related to the effect. That suggestion sounds peculiar to egalitarian ears: how could the administrator's taking a pen not have causally contributed *at all* or *in any sense* to the receptionist's problem? Though perhaps the suggestion is not so peculiar from a contextualist point of view; perhaps, for example, posing that very question induces a context that renders the alternative to the administrator's *not* taking a pen salient and therefore makes it true that her doing so *was* a cause.

A final possibility—at odds with much (but by no means all) of the literature on the metaphysics of causation—is that the ordinary concept of causation is scalar. In fact, it is fairly commonplace in x-phi to use a scale that explicitly asks participants about the 'strength' of a cause, or the 'extent to which' something is a cause. If the ordinary concept of causation is scalar, then it would be natural to interpret the non-extreme causal ratings given by many participants—both high and low—as indicating their judgements about the *extent to which* the various factors are causes. In that case, such responses would indicate neither uncertainty about whether the claim in question is true or false, nor a measure of the degree to which (or uncertainty about whether) the factor in question is contextually salient.[15]

A distinct but related proposal that is relevant to egalitarianism is that, while causation is *not* scalar, 'causal contribution' comes in degrees. Alex Kaiserman (2016) argues that causation is a *collective* rather than a *distributive* matter. It can be true that Alice and Bob surrounded the tree (by holding hands) but false that each of them individually surrounded the tree: their surrounding the tree is collective rather than distributive, and their contributions to the surrounding of the tree might be greater or lesser (Alice might contribute more by virtue of having longer arms, say). Similarly, he claims, '*x* and *z* caused *y*' should be read collectively. It is true neither that *x* caused *y* nor that *z* caused *y*; it is merely true of each of *x* and *z* that it was *among the causes* (or was *a* cause) of *y*. And, as with Alice and Bob, *x* and *z* can *contribute* to different degrees to the causing of *y.* On Kaiserman's view, then, causation is not scalar: it is infelicitous to say that *x* was more of a cause of *y* than was *z*, just as it is infelicitous to say that Alice surrounded the tree to a greater extent than did Bob (since

[15] See Woodward 2021: §7.2 for a discussion of causal strength ratings and their interpretation in experimental philosophy and psychology.

neither of them individually surrounded the tree *at all*). Given this kind of view, then, one might hold that non-extreme judgements in x-phi experiments are judgements about *degrees of causal contribution*. And in fact this delivers a further potential defensive move for the egalitarian. Even if *x* contributed more to the causing of *y* than did *z*, it is still true that each of *x* and *z* was among the causes of *y*—and equally so, since on this view it makes no sense to claim that *x* was *more* of a cause than was *z*.

Those philosophers who have claimed that causation (or causal contribution) comes in degrees have, however, generally avoided attempting to argue that scalarity (or degree of causal contribution) maps neatly onto causal selection (see e.g. Braham and Hees 2009; Moore 2009; Kaiserman 2016, 2017; Demirtas 2022). And indeed in those cases where scalar theories have actually been articulated, the determinants of degrees of causation or causal contribution have generally been taken to be objective factors as opposed to the kinds of broadly normative consideration that seem to drive causal selection. Insofar as this is so, such theories are in the spirit of (NR).

Halpern and Hitchcock (2015) are an exception to that general rule. Their 'graded' account of causation delivers, for example, the result that Professor Smith's taking the pen (*PT*=1) to be more of a cause of the problem (*PO*=1) than is the administrator's doing so (*AT*=1). Very roughly, the idea is that *PT*=1 is more of a cause of *PO*=1 than is *AT*=1 if the 'best witness' to *PT*=1 being a cause of *PO*=1 is more normal than is the best witness to *AT*=1 being a cause. The 'best witness' is the most normal possible world that demonstrates the counterfactual dependence of the effect on the cause. The most normal such world in the case of *PT*=1 is one where *PT*=0, *AT*=1 and *PO*=0, while the most normal such world in the case of *AT*=1 is one where *PT*=1, *AT*=0, and *PO*=0. Since *AT*=1 is more normal (in the sense of not violating any norms) than *PT*=1, the best witness to *PT*=1 causing the problem (a world where *AT*=1) is more normal than is the best witness to *AT*=1 (a world where *AT*=0); hence *PT*=1 is more of a cause of *PO*=1 than is *AT*=1. Halpern and Hitchcock's view constitutes a rejection of both (EC) and (NR), in that it builds normative considerations into the truth conditions for our ordinary causal claims.

The issue of the scalarity of the ordinary concept of causation (or causal contribution) is one that has not yet garnered very much explicit attention in the metaphysics literature (though see e.g. Bernstein 2017, Demirtas 2022, Kaiserman 2018, and Sartorio 2020 for recent discussions), and we shall not attempt to contribute to that nascent debate here. We merely note that the issue intersects in interesting ways both with experimental philosophy of causation in general and with the issue concerning egalitarianism in particular, and merits

further investigation. As things currently stand, it is unclear whether it is plausible to take non-extreme responses to x-phi questions to express judgements about degrees or 'strength' of causation or causal contribution, let alone whether or how doing so would impact on the debate about egalitarianism.

5. Conclusion

What we hope to have demonstrated here is that, as things stand, existing work on causation in experimental philosophy is far from conclusive on the question of whether or not the ordinary concept of causation is egalitarian, and hence does not get us any great traction on the metaphysical question of whether causation itself is a 'natural' relation. More significant progress will require closer attention to the kinds of the methodological issue discussed in §4, some of which—in particular the issue concerning what it is that the different degrees of 'agreement' with a given claim might really be measuring—are ones that apply much more broadly than to the causation debate.

Bibliography

Alicke, Mark. 1992. 'Culpable Causation'. *Journal of Personality and Social Psychology*, 63: 368–78.

— 2000. 'Culpable Control and the Psychology of Blame'. *Psychological Bulletin*, 1264: 556–74.

— and David Rose. 2012. 'Culpable Control and Causal Deviance'. *Social and Personality Psychology Compass*, 6: 723–35.

—, David Rose and Dori Bloom. 2011. 'Causation, Norm Violation and Culpable Control'. *The Journal of Philosophy*, 108: 670–96.

Bauer, Alexander Max and Jan Romann. 2022. 'Answers at Gunpoint: On Livengood and Sytsma's Revolver Case'. *Philosophy of Science*, 89: 180–92.

Bebb, Jon. 2022. 'Demarcating Contextualism and Contrastivism'. *Philosophy*, 97: 23–49.

Beebee, Helen. 2004. 'Causing and Nothingness'. In L.A. Paul, E.J. Hall and J. Collins. (eds), *Causation and Counterfactuals*. Cambridge, Mass.: MIT Press, 291–308.

— 2015. 'Causation, Projection, Inference, and Agency', in R.N. Johnson & M. Smith (eds), *Passions and Projections: Themes from the Philosophy of Simon Blackburn.* Oxford: Oxford University Press, 25-48.

Bernstein, Sara. 2017. ‘Causal Proportions and Moral Responsibility’. In D. Shoemaker (ed.), *Oxford Studies in Agency and Responsibility Volume 4*. Oxford.: Oxford University Press, 165–82.

Blanchard, Thomas and Schaffer, Jonathan. 2017. ‘Cause without Default’. In H. Beebee, C. Hitchcock, and H. Price. (eds), *Making a Difference: Essays on the Philosophy of Causation*. Oxford.: Oxford University Press, 175–214.

Bloom, Paul. 2007. ‘Religion is Natural’. *Developmental Science,* 10: 147–51.

Braham, Matthew, & Martin van Hees. 2009. ‘Degrees of Causation’. *Erkenntnis*, 71, 323–44.

Buckwalter, Wesley. m/s. ‘Pragmatics and Abnormal Causal Selection’.

Clarke, Randolph, Joshua Shepherd, John Stigall, Robin Repko Waller and Chris Zarpentine. 2015. ‘Causation, Norms, and Omissions: A Study of Causal Judgements’. *Philosophical Psychology*, 28: 279–93.

Danks, David, David Rose and Edouardo Machery. 2014. ‘Demoralizing Causation’. *Philosophical Studies*, 171: 1–27.

Demirtas, Huzeyfe. 2022. ‘Causation Comes in Degrees’. *Synthese*, 200: 64

Dowe, Phil. 2000. *Physical Causation*. Cambridge: Cambridge University Press.

— 2004. ‘Causes are Physically Connected to their Effects: Why Preventers and Omissions are not Causes’, in C. Hitchcock (ed.), *Contemporary Debates in Philosophy of Science*. Malden, MA: Blackwell, 189–96.

Gerstenberg, Tobias and Thomas Icard. 2020. ‘Expectations Affect Physical Causation Judgments’. *Journal of Experimental Psychology: General*, 149: 599–607.

Grice, H.P. 1975. ‘*Logic and Conversation’*, i*n* P. Cole & J.L. Morgan (eds), *Syntax and Semantics: Volume 3*. New York: Academic Press. 41–58.

Güver, Levin and Markus Kneer. 2023. ‘Causation and the Silly Norm Effect’. In S. Magen & K. Prochownik (eds), *Advances of Experimental Philosophy of Law*. London: Bloomsbury Press.

Halpern, Joseph Y. and Christopher Hitchcock. 2015. ‘Graded Causation and Defaults’. *The British Journal for the Philosophy of Science*, 66: 413–57.

Henne, Paul, Kevin O’Neill, Paul Bello, Sangeet Khemlani, and Felipe De Brigard. 2021. ‘Norms Affect Prospective Causal Judgements’. *Cognitive Science*, 44. DOI: 10.1111/cogs.12931

Hilton, Denis J.1990. ‘Conversational Processes and Causal Explanation’. *Psychological Bulletin*, 107: 65–81.

Hilton, Denis J. and Ben R. Slugoski. 1986. ‘Knowledge-Based Causal Attribution: The Abnormal Conditions Focus Model’. *Psychological Review*, 93: 75–88.

Hitchcock, Christopher. 1996. ‘Farewell to Binary Causation’. *Canadian Journal of Philosophy*, 26: 267–82.

— 2017. ‘Actual Causation: What’s the Use?’. In H. Beebee, C. Hitchcock, and H. Price. (eds), *Making a Difference: Essays on the Philosophy of Causation*. Oxford.: Oxford University Press, 116–31.

— and Knobe, Joshua. 2009. ‘Cause and Norm’. *Journal of Philosophy*, 106: 587–612.

Icard, Thomas, Jonathan F. Kominsky and Joshua Knobe. 2017. ‘Normality and Actual Causal Strength’. *Cognition*, 161: 80–93.

Kahneman, D. and D. T. Miller. 1986. ‘Norm Theory: Comparing Reality to its Alternatives’. *Psychological Review*, 93: 136–53.

Kaiserman, Alex. 2016. ‘Causal Contribution’. *Proceedings of the Aristotelian Society*, 116: 387–94.

— 2017. ‘Partial Liability’. *Legal Theory*, 23: 1–26.

— 2018. ‘ “More of a Cause”: Recent Work on Degrees of Causation and Responsibility’. *Philosophy Compass*, https://doi.org/10.1111/phc3.12498

Kirfel, Lara and David Lagnado. 2017. ‘Causal Judgments about Atypical Actions are Influenced by Agents’. *Cognition*, 212: Article 104721.

Knobe, Joshua. 2009. ‘Folk Judgments of Causation’. *Studies in History and Philosophy of Science*, 40: 238–42.

— and B. Fraser. 2008. ‘Causal Judgment and Moral Judgment: Two Experiments’. In W. Sinnott-Armstrong (ed.). *Moral Psychology: Volume 2*. Cambridge, MA: MIT Press. 441–8.

Kominsky, Jonathan F; Jonathan Phillips, Tobias Gerstenberg, David Lagnado and Joshua Knobe. 2015. ‘Causal Superseding’. *Cognition*, 137: 196–209.

Lewis, David K. 1973. ‘Causation’. *The Journal of Philosophy*, 70: 556–67.

— 1983. ‘New Work for a Theory of Universals’. *Australasian Journal of Philosophy,* 61: 343-77.

Livengood, Jonathan and Justin Sytsma. 2020. ‘Actual Causation and Compositionality’. *Philosophy of Science*, 87: 43–69.

—, Justin Sytsma and David Rose. 2017. ‘Following the FAD: Folk Attributions and Theories of Actual Causation’. *Review of Philosophy and Psychology*, 8: 273–94.

Maslen, Cei. 2004. ‘Causes, Contrasts, and the Nontransitivity of Causation’, in J. Collins, N. Hall & L.A. Paul (eds), *Causation and Counterfactuals*. Cambridge, MA: MIT Press, 341–57.

McGrath, Sarah. 2005. ‘Causation by Omission: A Dilemma’. *Philosophical Studies*, 123: 125–48.

Menzies, Peter. 2004. ‘Difference-making in Context’, in J. Collins, N. Hall & L.A. Paul (eds). *Causation and Counterfactuals*. Cambridge, MA: MIT Press, 138–80.

— 2006. ‘A Structural Equations Account of Negative Causation.’ http://philsciarchive.pitt.edu/2962 (deposited October 10, 2006).

— 2007. ‘Causation in Context’, in H. Price & R. Corry (eds), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford: Clarendon Press, 191–223.

Montminy, M. and A. Russo. 2016. ‘A Defense of Causal Invariantism’. *Analytic Philosophy*, 57: 49–75.

Moore, Michael. 2009. *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics*. Oxford: Oxford University Press.

Mumford, Stephen and Rani Lill Anjum. 2010. ‘A Powerful Theory of Causation’, in A. Marmodoro (ed.), *The Metaphysics of Powers*, London: Routledge, 143–59.

Northcott, Robert. 2008. ‘Causation and Contrast Classes’. *Philosophical Studies*, 139: 111–23.

Phillips, Jonathan, Jamie B. Luguri, and Joshua Knobe. 2015. ‘Unifying Morality’s Influence on Non-moral Judgments: The Relevance of Alternative Possibilities’. *Cognition,* 145: 30–42.

Price, Huw. 2007. ‘Causal Perspectivalism’, in H. Price & R. Corry (eds), *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*. Oxford: Oxford University Press, 250–92.

Putnam, H. 1982. ‘Why There Isn’t a Ready-Made World’. *Synthese*, 51: 141–57.

Ramsey, F. P. 1929. ‘General Propositions and Causality’, in his *Philosophical Papers*, ed. D. H. Mellor, Cambridge: Cambridge University Press, 145–63.

Rose, David. 2017. ‘Folk Intuitions of Actual Causation: A Two-pronged Debunking Explanation’. *Philosophical Studies*, 174: 1323–61.

Russell, Bertrand. 1912–13. On the Notion of Cause. *Proceedings of the Aristotelian Society* 13: 1–26.

Samland, Jana and Michael R. Waldman. 2015. ‘Highlighting the Causal Meaning of Causal Test Questions in Contexts of Norm Violations’, in D.C. Noelle, R. Dale, A.S.

Warlaumont, J. Yoshimi, T. Matlock, C.D. Jennings, & P.P. Maglio. (eds), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society. 2092–7.

— 2016. 'How Prescriptive Norms Influence Causal Inferences'. *Cognition*, 156: 164–76.

Samland, Jana, Marina Josephs, Michael R. Waldman and Hannes Rakoczy. 2016. 'The Role of Prescriptive Norms and Knowledge in Children's and Adults' Causal Selection'. *Journal of Experimental Psychology: General*, 145: 125–30.

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

— 2020. 'More of a Cause?'. *Journal of Applied Philosophy*, 37: 346–63.

Schaffer, Jonathan. 2000. 'Causation by Disconnection'. *Philosophy of Science*, 67: 285–300.

— 2005. 'Contrastive Causation'. *The Philosophical Review* 114: 297–328.

— 2013. 'Causal Contextualism'. In M. Blaauw (ed.), *Contrastivism in Philosophy: New Perspectives*. London: Routledge. 35–63.

Stapleton, Jane. 2008. 'Choosing What We Mean by "Causation in the Law" '. *Missouri Law Review,* 73(2): Article 6.

Strawson, P.F. 1992. 'Causation and Explanation', in his *Analysis and Metaphysics: An Introduction to Philosophy.* Oxford: Oxford University Press.

Sytsma, Justin. 2019. 'The Character of Causation: Investigating the Impact of Character, Knowledge, and Desire on Causal Attributions' (preprint). Available at http://philsciarchive.pitt.edu/16739/

— 2020. 'Causation, Responsibility, and Typicality'. *Review of Philosophy and Psychology*, 12: 699–719.

— 2022. 'The Responsibility Account'. In P. Willemsen and A. Wiegmann (eds), *Advances in Experimental Philosophy of Causation*. London: Bloomsbury Press. 145–64.

— and Jonathan Livengood. 2021. 'Causal Attributions and the Trolley Problem'. *Philosophical Psychology*, 34: 1167–91.

—, Jonathan Livengood and David Rose. 2012. 'Two Types of Typicality: Rethinking the Role of Statistical Typicality in Ordinary Causal Attributions'. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43: 814–20.

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

Wachbroit, Robert. 1994. ‘Normality as a Biological Concept’. *Philosophy of Science*, 61: 579–91.

Woodward, James. 2021. *Causation with a Human Face: Normative Theory and Descriptive Psychology.* New York: Oxford University Press.