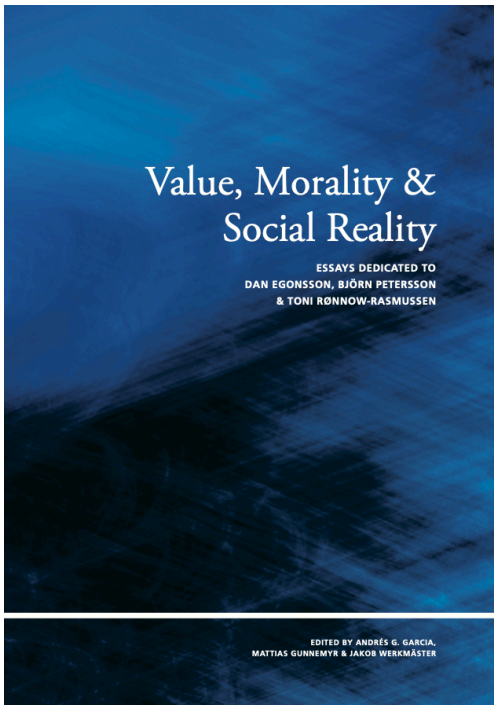
Causation, Responsibility, and Norms

Re-evaluating Our Norms in the Face of Climate Change

Caroline Torpe Touborg

In: Garcia, A., Gunnemyr, M. & Werkmäster, J. (2023) *Value, Morality & Social Reality: Essays dedicated to Dan Egonsson, Björn Petersson & Toni Rønnow-Rasmussen.* Lund: Department of Philosophy, Lund University. DOI: https://doi.org/10.37852/oblu.189

ISBN: 978-91-89415-65-2 (print), 978-91-89415-66-9 (digital)



Published by the Department of Philosophy, Lund University.
Edited by: Andrés Garcia, Mattias Gunnemyr, and Jakob Werkmäster
Cover image by Fabian Jones. Cover layout by Gunilla Albertén.

DOI: https://doi.org/10.37852/oblu.189.c539

This text is licensed under a Creative Commons Attribution-NonCommercial license.
This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license does not allow for commercial use.
(License: http://creativecommons.org/licenses/by-nc/4.0/)

Text © Andrés Garcia, Mattias Gunnemyr, and Jakob Werkmäster 2023.
Copyright of individual chapters is maintained by the chapters' authors.

Causation, Responsibility, and Norms

Re-evaluating Our Norms in the Face of Climate Change

Caroline Torpe Touborg

Abstract. In this paper, I combine two ideas. First: causation is relative to a possibility horizon, i.e., a class of possible worlds containing just those worlds that represent serious possibilities. Second: you are responsible for an outcome only if you have performed an action or omission that caused the outcome. Combining these two ideas raises a question: what is the relevant possibility horizon for evaluating whether the causal condition for responsibility is satisfied? I suggest that norms play a role in selecting the relevant possibility horizon. This yields a picture with norms as input. These norms select the relevant possibility horizon for assessing whether the causal condition for responsibility is satisfied. And together with other conditions for responsibility, this yields the output – namely, attributions of responsibility. This picture suggests a way to evaluate norms: when a bad outcome (such as climate change) occurs, taking one set of norms as input may yield the output that nobody is responsible – it just happened. By contrast, a different set of norms may allow us to attribute responsibility. In that case, I suggest that we should, *ceteris paribus*, prefer the norms that allow us to attribute responsibility for the outcome.

Until recently, it was widely assumed that causation is a binary relation: *c* is a cause of *e*. However, a growing number of authors now question this assumption.[1] An

[1] See footnote 3.

alternative suggestion is that causation is a three-place relation: *c* is a cause of *e* within a possibility horizon H, where a possibility horizon is simply a class of possible worlds containing just those possible worlds that represent serious possibilities. My aim in this paper is to explore what happens when we combine this idea with the widely held idea that causation is a necessary condition for responsibility: you are responsible for an outcome only if you have performed an action or omission that caused the outcome.

I begin by setting out the motivation for thinking that causation is a three-place relation with a possibility horizon as its third relatum (section 1). Once we combine this idea with the idea that causation is necessary for responsibility, we face a question: how do we select a possibility horizon for evaluating whether the causal condition for responsibility is satisfied? (section 2). In answering this question, my aim is to offer an idealised description of how we, as a matter of fact, make this selection. Based, among other things, on experimental work, we find that norms play a crucial role in how we select a possibility horizon. For moral responsibility, moral norms play this role; for legal responsibility, legal norms play this role, etc. (section 3). This yields a picture with norms as input; these norms select the possibility horizon we use to assess whether the causal condition for responsibility is satisfied; and together with other conditions for responsibility, this yields the output, which is attributions of responsibility (section 4).

In the second part of the paper, I argue that this picture of the relation between norms, causation, and responsibility suggests a way to evaluate norms: when a bad outcome (such as climate change) occurs, taking one set of norms as input may yield the output that nobody is responsible; taking a different set of norms as input may yield the output that some agents *are* responsible. In that case, I suggest that a system of norms that attributes responsibility for the bad outcome is, *ceteris paribus*, superior to a system of norms that does not: a system of norms that attributes responsibility for the bad outcome allows us to give a particular kind of *explanation* of why the bad outcome happened – namely, an explanation in terms of someone's having violated the norms (section 5). I end by showing how this way of evaluating norms may be used to re-evaluate our norms in the face of climate change (section 6).

There is an important shift between the first and the second part of the paper: in the first part of the paper, my aim is simply to offer a schematic *description* of how causation, responsibility, and norms relate to each other in our practice of attributing responsibility for outcomes. Such a description of our practice of course leaves it open whether this is how we *should* go about attributing responsibility for outcomes: is our practice of attributing responsibility for outcomes sound, or should it be revised? I do not attempt to answer this question. However, in the second part of the paper, I consider what follows if we assume that our practice *is* sound: I suggest that *if* our practice of attributing responsibility for outcomes is sound, then we may evaluate different systems of norms in terms of the attributions of responsibility they can support.

1. Causation Within a Possibility Horizon

In the following, I intend to talk about the ordinary, everyday notion of causation – the one we appeal to when we say that "the gust of wind caused the curtains to flutter" or "the lack of grass caused the wildebeest to migrate." This ordinary notion of causation has two important features. First, omissions and absences may play the role of cause and effect, as in the example above where the lack of grass (an absence) is cited as a cause of the wildebeest's migration. Second, causation is selective. As Hart and Honoré write:

> In most cases where a fire has broken out the lawyer, the historian, and the plain man would refuse to say that the cause of the fire was the presence of oxygen, though no fire would have occurred without it: they would reserve the title of cause for something of the order of a short-circuit, the dropping of a lighted cigarette, or lightning. (Hart and Honoré, 1985, p. 11)

This second feature of causation has prompted a growing number of authors to question the orthodoxy that causation is a binary relation between a cause *c* and an effect *e*. To see why, consider the following example:

Suppose that there is a lightning strike, and a forest fire starts thereafter. In ordinary contexts, such as a conversation among the forest rangers, it seems inappropriate to assert (1):

(1) The presence of oxygen caused the forest fire.

Indeed, if one of the forest rangers were to assert (1), it would be perfectly appropriate if a second forest ranger replied by denying this, saying

(2) The presence of oxygen did not cause the forest fire.[2]

There are, however, more unusual contexts where these judgements are reversed. Putnam gives the following charming example:

> Imagine that Venusians land on earth and observe a forest fire. One of them says, "*I* know what caused that – the atmosphere of the darned planet is saturated with oxygen." (Putnam, 1982, p. 150)

[2] In saying that it would be perfectly appropriate for a second forest ranger to respond by uttering (2), I am following the verdict of Kaiserman (2017), McGrath (2005), and Schaffer (2012). What matters here is simply that the second forest ranger *could* choose to respond by uttering (2), and this response would be appropriate. But of course, the second forest ranger *need* not respond in this way. She might instead respond in a more conciliatory way – first acknowledging that there is indeed a way to look at the situation where the presence of oxygen counts as a cause, and then directing attention towards the present causal inquiry: "why was there a fire now in this forest, rather than at some other time or in some other forest?" In *this* causal inquiry, the presence of oxygen is merely a background condition and does not count as a cause.

In this example, it is perfectly appropriate for the visiting Venusian to assert (1), and it would be inappropriate for the others to reply by asserting (2).

The orthodox view that causation is a binary relation has a hard time accommodating these data. Instead, several authors have suggested that causation has a third relatum, which we may call a possibility horizon. Let me explain this in more detail:

It is natural to understand the difference between the context of the forest rangers and the context of the visiting Venusians in terms of which possibilities are taken seriously in the two contexts. In the context of the forest rangers, the presence of oxygen is simply taken for granted – in this context, it is not treated as a serious possibility that there might have been no oxygen. In the context of the visiting Venusians, by contrast, it *is* treated as a serious possibility that there might have been no oxygen. We may use the notion of a *possibility horizon* to capture this idea, where a possibility horizon is simply a class of possible worlds. A possibility horizon H represents which possibilities we are taking seriously and which we are ignoring: if a world is included in H, it represents a possibility we are taking seriously; if a world is not included in H, it is being ignored. The suggestion then is that causation is relative to a possibility horizon:[3]

> *Ternary causation*: the causal relation has three relata and takes the form: *c* is a cause of *e* within possibility horizon H.

According to *Ternary causation*, the causal relation has three relata: the cause *c*, the effect *e*, and a possibility horizon H. Correspondingly, a complete causal claim takes the form "*c* is a cause of *e* within possibility horizon H." But of course, we do not usually mention possibility horizons when we are making causal claims. Rather, a typical causal claim takes the form "*c* is a cause of *e*," and the relevant possibility horizon is supplied by context. In a context where the relevant possibility horizon is H_1, for example, an utterance of "*c* is a cause of *e*" expresses the complete causal claim "*c* is a cause of *e* within possibility horizon H_1"; in a context where the relevant possibility horizon is H_2, an utterance of "*c* is a cause of *e*" expresses the different complete causal claim "*c* is a cause of *e* within possibility horizon H_2."

When is such a complete causal claim true? I will not here attempt to set out necessary and sufficient conditions for causation within a possibility horizon.[4] For

[3] This suggestion is implicit in the causal modelling approach to causation exemplified in the work of e.g. Halpern, Hitchcock, Pearl, and Woodward. According to the causal modelling approach, causation is relative to a causal model (see e.g. Halpern and Pearl, 2005, p. 845), and any given causal model represents certain possibilities and leaves out others. In this way, the claim that causation is relative to a model subsumes the claim that causation is relative to a possibility horizon. Woodward, for example, is clear that his causal modelling approach involves "relativizing causal judgments to a set of serious possibilities (or, what I take to be the same thing, to the choice of some system of representation that reflects those possibilities)" (Woodward, 2003, p. 90). Outside the causal modelling framework, the idea of relativizing causation to a possibility horizon is developed by Kaiserman (2017) and Touborg (2018).

[4] For an attempt to do so, see Touborg (2018).

now, all we need is the following *necessary* condition, which captures the idea that background conditions do not count as causes:

> *Candidate cause*: *c* is a cause of *e* within H only if there is a world in H where *c* does not occur.

In the case of the forest fire, this allows us to capture the data as follows: in the context of a conversation among the forest rangers, it is not treated as a serious possibility that there might have been no oxygen – the forest rangers are taking the presence of oxygen for granted. Correspondingly, the relevant possibility horizon $H_{Rangers}$ does not include any worlds where there is no oxygen. In the context of the forest rangers, (1) expresses the complete causal claim (1*):

> (1*) Within possibility horizon $H_{Rangers}$, the presence of oxygen caused the forest fire.

Since *Candidate cause* fails to be satisfied, (1*) is false. Correspondingly, its negation – which is expressed by (2) – comes out true. This fits with the observation that it is inappropriate to assert (1) in the context of the forest rangers and appropriate to assert (2).

In the context of the visiting Venusians, by contrast, it *is* treated as a serious possibility that there might have been no oxygen – after all, there is no oxygen where the Venusians come from. Correspondingly, the relevant possibility horizon $H_{Venusians}$ *does* include worlds where there is no oxygen. In the context of the Venusians, an utterance of (1) expresses the complete causal claim (1**):

> (1**) Within possibility horizon $H_{Venusians}$, the presence of oxygen caused the forest fire.

Here, *Candidate cause* is satisfied. It is also clear that the remaining conditions for causation (whatever they are) are satisfied. Thus, (1**) comes out true. Correspondingly, its negation – which is expressed by (2) – is false. This fits with the observation that it is appropriate to assert (1) in the context of the Venusians and inappropriate to assert (2).

2. Causation and Responsibility

How is causation related to responsibility?

The kind of responsibility I am interested in here is responsibility in the sense of *accountability*, as McKenna characterises it below:

> Treating another as accountable is treating her as one who is a candidate for moral demands and thus as one who is held to expectations that when complied with (or exceeded) merits praise and sometimes reward, and when violated merits blame and sometimes punishment. (McKenna, 2012, pp. 7-8)

Whenever I talk about "responsibility" in the remainder of this paper, it should be understood in this sense of accountability. The objects of responsibility – that is, the things one may be held responsible *for* – include actions, omissions, outcomes, and maybe more. In the following, I will focus on responsibility for outcomes – specifically, responsibility for *bad* outcomes.

It is widely held in the literature on moral responsibility that causation is a necessary condition for responsibility for outcomes.[5] This causal condition for responsibility is typically stated as follows:

> *The causal condition for responsibility*: you are responsible for a bad outcome *e* only if you have performed some action or omission *c*, such that *c* is a cause of *e*.

To see the intuitive support for *The causal condition for responsibility*, suppose that someone holds you responsible for some bad outcome *e* – say, the breaking of a vase. One way in which you might defend yourself is precisely by pointing out that you had nothing to do with the breaking of the vase: there is no causal connection between your behaviour and the breaking of the vase. If you succeed in showing that your behaviour did not cause the bad outcome, then it does indeed seem clear that you are not responsible for it.

The causal condition for responsibility simply requires that *c* is a cause of *e*. However, when we combine this with the idea that causation is a three-place relation with a possibility horizon as its third relatum, we face a crucial question: how do we (as a matter of describing our actual practice) select a possibility horizon for evaluating whether *The causal condition for responsibility* is satisfied?

3. How Do We Select a Possibility Horizon?

In this section, I am going to suggest that moral norms play a crucial role in our selection of a possibility horizon for evaluating whether *The causal condition for responsibility* is satisfied.

What are moral norms? As I shall use the term in the following, the content of a moral norm is simply a claim about what is morally permitted or required – for example, "you are morally required to keep your promises." Saying that a person is committed to a particular moral norm (or system of moral norms) amounts roughly to saying that she believes that these moral norms are true.

It has frequently been observed that our moral norms play a role in shaping our ordinary causal judgements, especially in contexts where we are concerned with attributing moral responsibility. Consider the following case:

[5] According to Sartorio (2007), this is the most widely held view about the relationship between causation and responsibility for outcomes. Against the causal condition for responsibility, see Sartorio (2004).

> *Flowers*: Billy has promised to water Suzy's flowers while she is away. However, he fails to do so, and the flowers die. If Billy had watered the flowers, they would have continued to bloom. It is also true that if the Queen had watered the flowers, they would have continued to bloom.[6]

In this case, it seems perfectly appropriate to assert (3) but inappropriate to assert (4):

(3) Billy's failure to water the flowers caused their death.

(4) The Queen's failure to water the flowers caused their death.

Furthermore, it seems clear that the reason we are treating Billy and the Queen differently in this case has to do with our moral norms: Billy's failure to water the flowers violated a moral norm that we are committed to (namely, the norm that "you are morally required to keep your promises"). By contrast, the Queen's failure to water the flowers did not violate any of our moral norms – according to our moral norms, she was under no obligation to look after Suzy's flowers.

The literature on causation as well as the experimental philosophy literature abounds with cases illustrating this phenomenon: when something bad happens (e.g. the flowers die),[7] we tend to treat behaviour that violates a moral norm we are committed to (e.g. Billy's failure to water the flowers as promised) as a candidate cause, while treating behaviour that is in accordance with our norms (e.g. the Queen's failure to water the flowers) as a mere background condition.[8] This suggests that norms play a crucial role in our selection of a possibility horizon: when someone violates a norm we are committed to, we take seriously the possibility that they might instead have acted as the norm requires. By contrast, when someone acts in accordance with the norms we are committed to, we do not take seriously the possibility that they might have acted differently.[9] In *Flowers*, for example, our

[6] This type of case has been widely discussed. See e.g. Hart and Honoré (1985), p. 38; Sartorio (2004); Beebee (2004); McGrath (2005); and Blanchard and Schaffer (2017). For experimental support for the verdicts I use, see Willemsen (2018) and Henne et al. (2017).

[7] The principle sketched here only seems to apply when we are asking about the causes of a bad outcome. For the sake of brevity, I omit discussion of neutral and good outcomes.

[8] For example, Henne et al. (2017) summarise their findings as follows (p. 274): "when an omission does not violate a norm (and there is counterfactual dependence) it will not be identified as a cause, and when it does violate a norm it will be identified as a cause." For an overview of the empirical literature, see Willemsen and Kirfel (2019); cf. Hitchcock and Knobe (2009).

[9] The question I am concerned with here is simply the question of which possibilities we take into consideration when we evaluate who caused a bad outcome, such as the flowers' death. It is a separate question whether an agent *could* have acted differently, in the sense that is relevant for free will. For example, it is perfectly consistent to say that the Queen *could* have acted differently (in the sense relevant for free will) – she could have made different choices about what to say, what to wear, etc.; perhaps, she could even have watered Suzy's flowers (if she had tried, she would have succeeded) – but we simply ignore these possibilities when we evaluate who is responsible for the flowers' death.

moral norms select a possibility horizon $H_{Flowers}$ that contains worlds in which Billy waters the flowers but does not contain worlds where the Queen waters the flowers. Relative to $H_{Flowers}$, it is then true that Billy's failure to water the flowers caused their death: *Candidate cause* is satisfied, and it seems clear that the remaining conditions for causation (whatever they are) are satisfied too. By contrast, it is false that the Queen's failure to water the flowers caused their death: since our possibility horizon does not include any world where the Queen waters the flowers, *Candidate cause* fails to be satisfied. This accommodates the intuitive judgement that it is appropriate to assert (3) but inappropriate to assert (4).

The way in which we evaluate causation for the purpose of attributing legal responsibility offers a clear parallel. While the relevant norms in the case of moral responsibility are moral norms, the relevant norms in the case of legal responsibility are specified by the laws and legal practice. Schaffer (2010) offers the following illustration: suppose that a lifeguard naps while he is on duty, and a swimmer drowns on his watch. Did the lifeguard's negligence cause the swimmer's death? When a judge is evaluating this question, she does not consider the closest possible world where the lifeguard does not nap – as Schaffer notes, that might be a world in which the lifeguard sneaks off to have a cigarette, and what happens in such a world seems clearly irrelevant to the question we are interested in. Rather, the judge considers the closest world in which the lifeguard lives up to what is legally required of him. Based on this and other cases, Schaffer suggests the following description of our current legal practice:[10]

> Generalizing, it seems that causal judgments in the law are based on a comparison between the actual course of events and an alternative scenario in which the defendant acts lawfully. (Schaffer, 2010, p. 272)

Similarly, Halpern and Hitchcock make the following observations about the role of legal norms in determining legal causation:

> The law suggests a variety of principles for determining the norms that are used in the evaluation of actual causation. In criminal law, norms are determined by direct legislation. For example, if there are legal standards for the strength of seat belts in an automobile, a seat belt that did not meet this standard could be judged a cause of a traffic fatality. By contrast, if a seat belt complied with the legal standard, but nonetheless broke because of the extreme forces it was subjected to during a

[10] Schaffer uses his contrastive account of causation (Schaffer, 2005) to accommodate this observation: according to Schaffer, causation is a four-place relation between a cause, a contrast to the cause, an effect, and a contrast to the effect. When we are evaluating causation in the law, Schaffer's suggestion then is that the relevant contrast to the cause is lawful conduct. However, Schaffer's more general observation – namely, that when we are evaluating causation in the law, we take seriously the possibility that the defendant might have acted in accordance with the law, but do not take seriously e.g. the possibility that the defendant might have broken the law in some other way (say, by sneaking off to get a cigarette) – may just as easily be understood as an observation about how we select a possibility horizon.

> particular accident, the fatality would be blamed on the circumstances of the accident, rather than the seat belt. In such a case, the manufacturers of the seat belt would not be guilty of criminal negligence. In contract law, compliance with the terms of a contract has the force of a norm. In tort law, actions are often judged against the standard of "the reasonable person." (Halpern and Hitchcock, 2010)

We may think of the law as guiding our selection of which possibilities we do and do not take seriously when we are evaluating legal causation: the possibility horizon we select for the purpose of evaluating causation in the law is determined, to a large extent, by the content of the law. We select a possibility horizon that includes possible worlds where the defendant acts in accordance with the law (more carefully, where the defendant does what the law minimally requires), while it leaves out possible worlds where the defendant breaks the law in some other way than he actually did, or where people who in fact acted in accordance with the law act differently.

Let us take stock. The picture that has emerged so far suggests that there are different *flavours* of responsibility: when we are concerned with moral responsibility, we assess whether *The causal condition for responsibility* is satisfied by considering causation within the possibility horizon that is selected by our moral norms (that is, the moral norms we believe to be true). When we are concerned with legal responsibility (within a particular jurisdiction), we assess whether *The causal condition for responsibility* is satisfied by considering causation within the possibility horizon that is selected by the relevant legal norms (that is, the legal norms that are contained in the laws and legal practice of the relevant jurisdiction). And there may be additional flavours of responsibility following the same pattern. For example, there might be a particular flavour of responsibility that applies to sports. Here, the relevant norms might be norms of skill – for example, the level of skill that can be expected of players in the Champions League – and a coach or sports journalist might hold a player responsible for a failure to score a goal when their actions or omissions count as causes relative to the possibility horizon selected by the relevant norms of skill.[11] In the following, I will leave these other flavours aside and focus on moral responsibility.

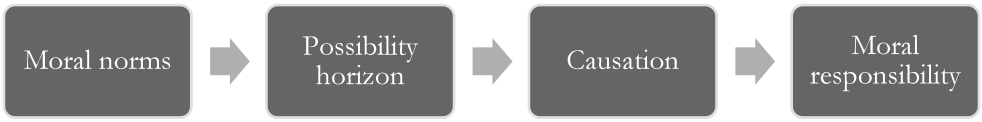
4. The Emerging Picture

We have now arrived at a picture where our moral norms play a role in how we select a possibility horizon when the purpose of our causal inquiry is to attribute moral responsibility for some bad outcome.[12] This possibility horizon in turn shapes

[11] See e.g. the general account of blame and credit in Björnsson (2017).

[12] In a context where we have a different purpose – for example, understanding the physics of how forest fires occur – it may well be the case that our moral or legal norms play no role in our selection of a possibility horizon.

our causal judgements; and these causal judgements in turn inform our attributions of responsibility. These relations between moral norms, possibility horizons, causation, and moral responsibility are summarized in the diagram below:



One way in which we may put this picture to use is in understanding disagreements: the picture suggests that disagreements about moral responsibility may be the result of underlying disagreements about causation, which are themselves the result of underlying disagreements about moral norms. More carefully, suppose that A takes S_A to be the correct system of moral norms and B takes a different system S_B to be the correct system of moral norms. In this situation, A is going to select a possibility horizon H_A on the basis of S_A and use H_A as the basis for her causal judgements and eventually her attributions of responsibility. Similarly, B is going to select a possibility horizon H_B on the basis of S_B and use H_B as the basis for her causal judgements and eventually her attributions of responsibility. Thus, A and B may reach different verdicts about causation and moral responsibility because of an underlying disagreement about whether S_A or S_B is the correct system of moral norms.[13]

This picture resembles a suggestion made by Mackie in his paper "Responsibility and language" (1955). Mackie here presents the following case:

> In Sydney some time ago a motor cyclist was exceeding the speed limit; a traffic policeman, also on a motor cycle, chased him, and soon they were both travelling, according to the reports, at 70 m.p.h. Then an unobservant citizen stepped off a bus into the policeman's path; in the crash that resulted the other man was killed at once; the policeman died the next day.
>
> There was some disagreement as to who was responsible for this accident. The police announced that when they caught the original speedster they would charge him with causing the two deaths. The general public was inclined at first to hold the policeman responsible for the other man's death, but tended to change its mind a little when he died himself. (Mackie, 1955, p. 143)

This disagreement about responsibility, Mackie suggests, is the result of an underlying disagreement about what is "the normal, proper, or expected course of events":

[13] For a nice example of how disagreements about moral norms may lead to disagreements about causation, see Kaiserman (2017), pp. 57-58.

> The answer we choose will depend on what we take to be the normal, proper, or expected course of events; the person that we hold responsible is the one who steps outside this expected pattern. Thus if we assume, as apparently the police did, that it is normal and proper for traffic police to pursue, relentlessly and with all the means in their power, those who break the speed limit, and that it is normal and proper for people to step off buses without taking precautions against motor cycles passing at 70 m.p.h., but that it is not normal and proper for cyclists (other than police in pursuit of a criminal) to break the speed limit, then we shall hold the cyclist responsible. The behaviour of the policeman and the bus traveller belonged, on this view, to the normal pattern, but that of the cyclist was an intrusion into it.
>
> On the other hand, the general public is inclined to take a less legalistic view, and not to identify what is normal and proper with strict conformity to the law and the police regulations. It might hold, therefore, that the cyclist's conduct, though illegal, was yet normal and expected, including his increase in speed when chased, whereas it was not normal, not "reasonable", for the policeman to go to such lengths to catch a speedster. Making these assumptions, the general public would conclude that the policeman was responsible for the accident. (Mackie, 1955, p. 144)

The structure of Mackie's suggestion is parallel to what I have suggested above: in both cases, the idea is that we may trace disagreements about moral responsibility back to disagreements about causation, which may in turn be traced back to an underlying disagreement about what should be treated as a mere background condition (as belonging to "the normal pattern" in Mackie's terms) and what should be treated as a candidate cause (an "intrusion" into the normal pattern). The main difference between my suggestion and Mackie's consists in the fact that Mackie traces disagreements about what should be treated as a mere background condition and what should be treated as a candidate cause to disagreements about what is "the normal, proper, or expected course of events," where what is "expected" is understood broadly, including both what is expected in a normative sense and what is statistically expected. By contrast, I trace disagreements about moral responsibility more specifically to disagreements about moral norms.

From here, there are several ways one might go. So far, I have focused on *describing* the relation between causation, responsibility, and moral norms in our practice of attributing moral responsibility for outcomes. One question one might consider at this point is whether this practice is sound. For now, I will put this question aside. In the following, I will instead consider what happens if we accept the picture I have set out above, not only as a descriptive picture of how we in fact go about attributing responsibility for outcomes, but also as a prescriptive picture of how we *should* go about attributing responsibility for outcomes. I will argue that *if* we accept that this is how we should go about attributing responsibility for outcomes, the picture I have sketched suggests a way to *evaluate* different systems of norms in terms of the attributions of responsibility they can and cannot support.

5. Evaluating Systems of Norms

The picture presented in sections 1-4 sets out how taking a particular system of moral norms as input yields particular attributions of moral responsibility as output. This imposes a requirement of consistency between our moral norms and our attributions of responsibility: the verdicts we get when we use our moral norms as input have to *support* the attributions of responsibility we make. In our everyday lives, we typically achieve consistency by taking our system of moral norms as given and then simply accepting the verdicts about moral responsibility that it delivers. However, we may also look at things the other way around: if we have some prior understanding of the verdicts about responsibility we *should* be getting – and in this section, I will suggest that there are independent reasons to prefer some verdicts over others – then we may *reverse-engineer* our system of moral norms so that it can support those verdicts. To the extent that we know what the output (attributions of responsibility) should be, the picture therefore offers a way to compare and evaluate different systems of norms in terms of their ability to support those verdicts.

Let me illustrate this with a simple example: suppose that Ben starts a camp fire in a dry forest, a gust of wind blows a spark into some nearby dry grass, the grass catches fire, and a large and destructive forest fire develops. Is Ben responsible for the forest fire? The answer depends on which norms we take as input. Consider the following two moral norms:

N_1: It is not permissible to start a camp fire in a dry forest.

N_2: It is permissible to start a camp fire in a dry forest.

Using N_1 as input, we get a possibility horizon H_1 that includes worlds where Ben does not start a camp fire. Relative to H_1, we find that Ben's action is a cause of the forest fire: *Candidate cause* is satisfied and the remaining conditions for causation (whatever they are) are clearly satisfied too. Thus, Ben satisfies *The causal condition for responsibility*. Of course, *The causal condition for responsibility* is not a sufficient condition for moral responsibility. However, provided that Ben also satisfies the remaining conditions, we get the verdict that he is responsible for the forest fire.[14]

Using N_2 as input, on the other hand, we get a possibility horizon H_2 where Ben starts a camp fire in *every* world in H_2. That is, Ben's starting a camp fire is simply

[14] What are these other conditions for being responsible for an outcome? I will not attempt a complete answer to this question here. However, at least the following requirement seems plausible: an agent is only responsible for an outcome if he is responsible for the action or omission that in turn caused the outcome. For example, Ben is only responsible for the forest fire if he is responsible for starting the camp fire: we should not hold him responsible for the forest fire if someone held a gun to his head and forced him to start the camp fire. For an overview, see e.g. Sartorio (2007).

treated as a background condition in H_2. Relative to H_2, we find that Ben's action is *not* a cause of the forest fire: *Candidate cause* fails to be satisfied. Thus, Ben does not satisfy *The causal condition for responsibility*, and it immediately follows that he is not responsible for the forest fire. If we also treat the other circumstances – the gust of wind, the proximity of the dry grass, and the dryness of the forest – as background conditions, we do not find *any* causes of the forest fire. Instead, the forest fire appears inevitable: it occurs in every world in H_2. If we instead treat it as a serious possibility that there might not have been a gust of wind, that there might not have been dry grass where the spark landed, or that the forest might not have been so dry, we do find causes of the forest fire – namely, the gust of wind, the proximity of the dry grass, and the dryness of the forest. But either way, Ben's action does not count as a cause of the forest fire, and so we get the verdict that he is not responsible.[15]

This reveals an important difference between the two norms, N_1 and N_2: N_1 allows us to hold an agent responsible for the forest fire – namely Ben, who should have refrained from starting his camp fire. By contrast, N_2 does not allow us to hold anyone responsible.

Do we have a principled reason to prefer one of these verdicts over the other? I believe we do: as a tentative suggestion, one might think of a system of moral norms as a theory for explaining bad outcomes in terms of the actions and omissions of moral agents. One such theory is superior to another when it can explain more. For example, considered as a theory for explaining bad outcomes in terms of the actions and omissions of moral agents, a system of norms that includes N_1 is, *ceteris paribus*, superior to a system of norms that includes N_2 instead: a system of norms that includes N_1 can *explain* the occurrence of the forest fire (a bad outcome) in terms of Ben's lighting a camp fire; by contrast, N_2 cannot offer any explanation of this outcome in terms of the actions and omissions of moral agents – according to N_2, the forest fire is simply an accident (either seen as inevitable or as being caused by non-agential features of the situation, such as the gust of wind, the proximity of dry grass, or the dryness of the forest).

[15] A person who is committed to N_2 may still recognize that there is a perspective from which Ben's starting the camp fire *is* a cause of the forest fire: for example, when the purpose of the causal inquiry is to understand the *physics* of how the forest fire started, moral norms play no role in the selection of a possibility horizon – and thus, even someone who is committed to N_2 may select a possibility horizon $H_{Physics}$ that treats it as a serious possibility that Ben might not have started a camp fire. The important point, however, is that, to a person who is committed to N_2, the fact that Ben's starting a camp fire is a cause of the forest fire within $H_{Physics}$ has no bearing on the question whether *The causal condition for responsibility* is satisfied: what matters for this question is whether Ben's starting the camp fire is a cause relative to the relevant *normative* possibility horizon. A different example might bring out the idea more clearly: suppose you decide to walk home one evening instead of taking a cab. On the way home, you get robbed. In this case, there obviously is a perspective from which your decision to walk home is a cause of your getting robbed – if you had taken a cab instead, you would not have been robbed. However, in a discussion about who is responsible for the robbery, it would seem entirely misplaced to point out that your decision to walk home was a cause of your getting robbed: the norm-free perspective from which your decision to walk home is a cause of the robbery is simply irrelevant when we want to determine whether *The causal condition for responsibility* is satisfied.

The suggestion that we may think of a system of moral norms as a theory for explaining bad outcomes in terms of the actions and omissions of moral agents draws on the principle of moral harmony. In *Utilitarianism and Co-operation*, Regan expresses this principle as follows:

> [T]here is the intuition that whatever the correct moral theory is, […] [i]t ought to be the case that if all agents satisfy the theory, then the class of all agents produce the best consequences they can produce collectively by any pattern of behaviour. (Regan, 1980, p. 3)

By contraposition, the principle states that if a suboptimal outcome is produced, there must be at least one agent who has failed to satisfy the moral theory, i.e. who has failed to behave as the theory requires (Pinkert, 2015, pp. 975-77). It is natural to understand this precisely as a principle about explanation: when a suboptimal outcome occurs, the correct moral theory will explain *why* it occurred in terms of one or more agents failing to behave as the theory requires.

This gives us a way to evaluate different systems of norms: a system of norms that allows us to explain a bad outcome in terms of the actions and omissions of moral agents is, *ceteris paribus*, superior to an alternative system of norms that treats the outcome as unexplainable – i.e., as something that "just happened." This is, of course, not the only relevant consideration when evaluating systems of norms: sometimes a system of norms that allows us to explain a bad outcome in terms of the actions and omissions of moral agents should still be rejected because it is deficient in other ways. However, when two systems of norms are equally good in all other respects, I believe we should prefer the one with more explanatory power.

6. Evaluating Our Norms in the Face of Climate Change

In a newspaper article about the deadly 2021 heatwave in Canada, the climate journalist Eric Holthaus wrote:

> Climate change isn't just a thing that's happening, it's a series of choices made by actual people who are sharing this planet with us. (Holthaus, 2021)

The picture I have sketched here yields an intriguing interpretation of this statement:[16]

First, there exist some systems of norms according to which climate change, and more specifically climate-change-related disasters, "just happen" – that is, these systems of norms do not allow us to *explain* these disasters in terms of the actions and omissions of moral agents. Consider, for example, the following simplified system of norms:

[16] This interpretation may, of course, go beyond what Holthaus intended; that is not the point.

Unrestricted permission to emit greenhouse gases

The norms for individuals, businesses, and governments are as follows:

a) *Individuals:* it is permissible to drive, fly, shop, eat an ordinary Western diet, and do all the other things that are part of an affluent Western lifestyle.
b) *Businesses:* within legal limits, it is permissible to produce any goods that are demanded in the market, in whichever way is most profitable.
c) *Governments:* it is permissible for members of the public to vote and prioritise political issues as they see fit, without regard for the interests of future generations; and it is permissible for governments to act in accordance with the priorities of the general public.

If we use this system of norms as input, we find that climate-change-related disasters "just happen" – they are not caused by the actions and omissions of moral agents, and thus, we cannot appropriately hold anyone responsible for them.

Second, there also exist systems of norms which would not commit us to the conclusion that climate change is "just a thing that's happening." Consider, for example, the simplified system of norms below:

Restricted permission to emit greenhouse gases

The norms for individuals, businesses, and governments are as follows:

a*) *Individuals:* it is not permissible to emit greenhouse gases for the sake of trivial benefits.
b*) *Businesses:* it is not permissible to emit greenhouse gases during the production of goods and services, when these emissions could easily be avoided.
c*) *Governments:* it is not permissible for voters and governments to neglect issues that could be devastating for future generations.

The precise content of these norms is not important. The important point is that there exist systems of norms – exemplified by *Restricted permission* – that would allow us to hold individuals, businesses, and governments responsible for climate-change-related disasters. If we were to adopt such a system of norms, we would indeed see climate change as the result of "a series of choices made by actual people who are sharing this planet with us."

By stating that "[c]limate change isn't just a thing that's happening, it's a series of choices made by actual people who are sharing this planet with us," Holthaus implicitly rejects norms such as *Unrestricted permission*, according to which climate change "just happens," and endorses norms such as *Restricted permission*, according to which climate change is indeed the result of "a series of choices made by actual people who are sharing this planet with us." The arguments I have presented support this: considered as a theory for explaining bad outcomes in terms of the actions and omissions of moral agents, a system of norms that includes *Restricted permission* is, *ceteris paribus*, superior to a system of norms that includes *Unrestricted permission*.

7. Conclusion

In this paper I have done two things. First, I have suggested a picture of the relation between moral norms, causation, and moral responsibility: our moral norms play a role in our selection of a possibility horizon for assessing whether the causal condition for moral responsibility is satisfied. Second, I have suggested that, if we assume that this picture captures how we *should* attribute responsibility for outcomes, we get a way to evaluate different systems of moral norms: when a bad outcome happens, a system of norms that allows us to *explain* this bad outcome in terms of the actions and omissions of moral agents is, *ceteris paribus*, superior to a system of norms that treats the outcome as something that "just happens." Applied to climate change, this suggests that norms such as *Restricted permission* are, *ceteris paribus*, preferable to norms such as *Unrestricted permission*.[17]

References

Beebee, Helen (2004) "Causing and nothingness" in J. Collins, N. Hall, & L. A. Paul (Eds.) *Causation and counterfactuals* (291-308). Cambridge, Mass.: MIT Press.

Björnsson, Gunnar (2017) "Explaining (away) the epistemic condition on moral responsibility" in P. Robichaud, & J. W. Wieland (Eds.) *Responsibility: the epistemic condition* (146-62). Oxford: Oxford University Press.

Blanchard, Thomas, & Jonathan Schaffer (2017) "Cause without default" in H. Beebee, C. Hitchcock and H. Price (Eds.) *Making a difference: essays on the philosophy of causation* (175-214). Oxford: Oxford University Press.

Halpern, Joseph Y., & Judea Pearl (2005) "Causes and explanations: a structural-model approach. Part I: causes". *British Journal for the Philosophy of Science*, *56*(4): 843-887.

Halpern, Joseph Y., & Christopher Hitchcock (2010) "Actual causation and the art of modelling" in R. Dechter, H. Geffner, & J. Halpern (Eds.) *Causality, probability, and heuristics: a tribute to Judea Pearl* (383-406). London: College Publications.

Hart, H. L. A., & Tony Honoré (1985) *Causation in the law*. Oxford: Clarendon Press.

Henne, Paul, Ángel Pinillos, & Felipe De Brigard (2017) "Cause by omission and norm: not watering plants". *Australasian Journal of Philosophy*, *95*(2): 270-283.

Hitchcock, Christopher, & Joshua Knobe (2009) "Cause and norm". *Journal of Philosophy*, *106*(11): 587-612.

[17] I am grateful to audiences at the Lund Higher Seminar in Philosophy, the Umeå Higher Seminar in Philosophy, the "Causation and Responsibility" workshop at Bern University, and the Stockholm June Workshop in Philosophy. I would especially like to thank Per Algander, Erik Carlson, Anton Emilsson, Mattias Gunnemyr, Madeleine Hayenhjelm, Sofia Jeppsson, Alex Kaiserman, Christian Löw, Luise Mirow, Matt Talbert, Bram Vaassen, and Martín Abreu Zavaleta for their detailed comments.

Holthaus, Eric (30th June 2021) "How did a small town in Canada become one of the hottest places on Earth?" *The Guardian*. https://www.theguardian.com/commentisfree/2021/jun/30/lytton-hottest-places-world-climate-emergency

Kaiserman, Alex (2017) "Necessary connections in context". *Erkenntnis*, *82*(1): 45-64.

Mackie, John L. (1955) "Responsibility and language". *Australasian Journal of Philosophy*, *33*(3): 143-159.

McGrath, Sarah (2005) "Causation by omission: a dilemma". *Philosophical Studies*, *123*(1-2): 125-148.

McKenna, Michael (2012) *Conversation and responsibility*. Oxford: Oxford University Press.

Pinkert, Felix (2015) "What if I cannot make a difference (and know it)". *Ethics*, *125*(4): 971-998.

Putnam, Hilary (1982) "Why there isn't a ready-made world". *Synthese*, *51*(2): 141-167.

Regan, Donald (1980) *Utilitarianism and co-operation*. Oxford: Clarendon Press.

Sartorio, Carolina (2004) "How to be responsible for something without causing it". *Philosophical Perspectives*, *18*(1): 315-36.

Sartorio, Carolina (2007) "Causation and responsibility". *Philosophy Compass*, *2*(5): 749-765.

Schaffer, Jonathan (2005) "Contrastive causation". *The Philosophical Review*, *114*(3): 327-58.

Schaffer, Jonathan (2010) "Contrastive causation in the law". *Legal Theory*, *16*(4): 259-97.

Schaffer, Jonathan (2012) "Causal contextualism" in M. Blaauw (Ed.) *Contrastivism in philosophy* (35-63). New York: Routledge.

Touborg, Caroline (2018) *The dual nature of causation: two necessary and jointly sufficient conditions* (Doctoral thesis). University of St Andrews. https://research-repository.st-andrews.ac.uk/handle/10023/16561

Willemsen, Pascale (2018) "Omissions and expectations: a new approach to the things we failed to do". *Synthese*, *195*(4): 1587-1614.

Willemsen, Pascale, & Lara Kirfel (2019) "Recent empirical work on the relationship between causal judgements and norms". *Philosophy Compass*, *14*(1): e12562.

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