PART II

Dispositions and Causes

5

A disposition-based process-theory of causation

Andreas Hüttemann

Given certain well-known observations by Mach and Russell, the question arises what place there is for causation in the physical world. My aim in this chapter is to understand under what conditions we can use causal terminology and how it fits in with what physics has to say. I will argue for a disposition-based process-theory of causation. After addressing Mach's and Russell's concerns I will start by outlining the kind of problem the disposition-based process-theory of causation is meant to solve. In a second step I will discuss the nature of those dispositions that will be relevant for our question. In section 3 I will discuss existing dispositional accounts of causation before I proceed to present my own account (sections 4 to 6) and contrast it with traditional process-theories (section 7).

1 Problems of causation

There are many problems of causation. Michael Tooley and Ernest Sosa for instance present a list of five fundamental issues: (1) the relation of causal laws and causal relations; (2) do causal states of affairs logically supervene on non-causal ones? (3) if not, is an a posteriori reduction feasible? (4) are causal relations immediately given in experience? (5) do causal concepts need to be analysed, or can they be taken to be basic? (Sosa and Tooley 1993: 5). While all of these are important issues, I will not attempt to answer any of these—at least not directly. Rather, I will approach the issue of causation from a different angle. In what follows I will focus on 'probably the central problem in the metaphysics of causation' (Field 2003: 443).

At the turn of the 20th century Mach and Russell suggested that the terms 'cause' and 'effect' ought to be eliminated because they are imprecise and their conditions of application inconsistent with what physics says about reality. However, causal notions play an important role in everyday life and in the special sciences, and, as Nancy Cartwright (1983: 21ff) has pointed out, they seem indispensable in order to distinguish efficient from non-efficient strategies. How are these observations to be reconciled? What place have causes and effects in a world as described by physics—for short: in a physical world?

1.1 Historical background

In 1874 Gustav Kirchhoff criticized the notion of causation in the well-known preface to his *Lectures on Mechanics*:

It is common to define mechanics as the science of forces and forces as causes that bring about motions or tend to bring them about. [...]. [This definition] is tainted by an obscurity that is due to the notion of cause and tendency [...]. For this reason I propose that the aim of mechanics is to describe the movements that take place in nature – more specifically to describe them completely and as simple as possible. What I want to say is that we should aim at stating what phenomena there are rather than to determine their causes. (Kirchhoff 1874, preface)

Kirchhoff's criticism concerns causation in a productive sense or understood as a force, which was indeed the main concept of cause that was used by physicists in the first half of the 19th century.[1] Eliminating productive causes from physics still leaves room for other conceptions. Gustav Theodor Fechner and the early Ernst Mach for example turned to John Stuart Mill. Mill had repudiated productive causes and defended a regularity view according to which the cause is an instance of the antecedent in a law, which is a sufficient condition for the occurrence of the effect:

To certain facts certain facts always do, and, as we believe, will continue to, succeed. The invariable antecedent is termed the cause; the invariable consequent the effect. (Mill 1843/1891: 213)

When at the turn of the century Mach and Bertrand Russell criticized the notion of cause it was a Millian regularity view of causation they had in mind.

Mach argued that for at least three reasons the concept of cause cannot be applied to reality as described by physics and should therefore be given up.

(i) Taking seriously the concept of cause as a set of conditions implies that you need to consider every single factor on which an event depends. That is practically impossible.

If one attempts to eliminate the traces of fetishism, which are still associated with the concept of cause, and takes into consideration that in general you cannot specify a cause since a phenomenon most of the times is determined by a whole system of conditions, you are led to the conclusion to give up the concept of cause altogether. (Mach 1896: 435–6)

(ii) The concept of cause requires strict regularities. However, there are no such regularities:

In nature there are no causes and no effects. Nature is there only once. Repetitions of identical cases, such that A is always correlated with B, the same outcome under the same conditions, i.e. what's essential for the relation between cause and effect, exists only in abstraction...(Mach 1883: 459)

[1] See Hüttemann, 'The Elimination of Causal vocabulary from Physics', manuscript.

(iii) Finally, the advanced sciences replace causal terminology by the concept of a mathematical function, which is a more precise notion.

In the higher developed natural sciences the use of the concepts cause and effect becomes more and more limited. There is a perfectly good reason for this, namely that these concepts describe a state of affairs provisionally and incompletely – they are imprecise [...]. As soon as it is possible to characterise the elements of events through measurable quantities, [...] the dependence of the elements on each other can be characterized more completely and more precisely through the concept of a function, rather than through the insufficiently determined concepts cause and effect. (Mach 1883, 278)

Russell in his well-known paper 'On the notion of Cause' adds two important considerations:

(iv) Causes are usually conceived as localized events (locality). However, no localized event is sufficient for the occurrence of any other event, because there may always be an interfering factor.

In order to be sure of the expected effect, we must know that there is nothing in the environment to interfere with it. But this means that the supposed cause is not, by itself, adequate to insure the effect. And as soon as we include the environment, the probability of repetition is diminished, until at last, when the whole environment is included, the probability becomes *nil*. (Russell 1912–13, 7–8)

Thus, if the fact that causes determine their effects is spelled out in terms of conditional regularities such that the cause is the antecedent, we cannot have both locality and determination.

(v) In a physical system as described by fundamental physics, the future determines the past in exactly the same way as vice versa.

...the future 'determines' the past in exactly the same sense in which the past 'determines' the future. The word 'determine', here, has a purely logical significance: a certain number of variables 'determine' another variable if that variable is a function of them. (Russell 1912–13: 15)

Therefore, the asymmetry which we associate with the causal relation is inconsistent with fundamental physics.

Mach and Russell both conclude that there is no place for causation in the advanced sciences. Even though this latter claim has been proven false for today's advanced sciences,[2] the issue has been raised of how to reconcile Mach's and Russell's observations with the persistence and usefulness of causal terminology.

[2] See, for instance (Suppes 1970: 5–6), (Hitchcock 2007: 55), and (Williamson 2009: 195–7).

1.2 The folk-conception of causation

As others have observed (Norton 2007), there is probably no univocal conception of causation as we usually (pre-theoretically) understand it, let alone a conception that furthermore applies to the special sciences. What people mean by causation has probably changed over time and even within historical periods of time there might be different conceptions. Be that as it may, I will present some features that are often associated with causation in everyday life as well as in the special sciences, and will ask to what extent a relation that has these features has its place in a physical world.

- *Modal Force*: A cause brings about the effect. It somehow *forces* the effect to occur. The cause determines the effect to occur.
- *Asymmetry*: Causes bring about effects, but not vice versa.
- *Time-Precedence*: Causes precede their effects.
- *Locality*: Causes and effects are events that can be localized in spacetime, in some definite kind of region.
- *Dominant Cause*: There is a distinction between a main cause and secondary factors (between causes and conditions).
- *Objectivity*: Whether or not something is a cause is independent of human interests or convictions.
- *Multi-level*: Causal relations obtain on all kinds of 'levels': in physics, the special sciences in everyday life.[3]

As we have seen, some of these characteristics are somewhat problematic in the light of the early 20th-century criticism. In what follows, I will focus on the problem of modal force, and occasionally mention how I think the other features might be dealt with. I will not have anything to say about the problem of the asymmetry of the causal relation, but assume that an account can be given that explains this feature in terms of statistical mechanics somewhere along the lines of Albert (2000) or Field (2003, section 4).

In what follows I will argue that, even though causation in the above-mentioned sense probably has no place in a physical world, a near relative can be shown to be compatible with physics—at least given the right conditions. That will allow us to understand why and under what conditions we are successfully applying causal terminology. I take it that the relation between causes and effects does not describe *sui generis* facts over and above those described by the various sciences. Causal facts—I will argue—are partly determined by the facts of physics, biology, economics, etc., but they are also partly determined by pragmatic considerations.

[3] Some of these characteristics are mentioned in (Norton 2007). *Modal Force* and *Asymmetry* have been called the hard problems of a theory of causation (Hitchcock 2007: 58).

2 Laws and dispositions

One of the central claims I will argue for is that we need to understand the roles of dispositions for physical laws. This then allows us to understand how causal relations emerge from compound systems given the right kind of conditions (as well as certain pragmatic considerations that I will talk about towards the end).

Law, disposition, and causation are closely related concepts. All three are concerned with some kind of natural necessity. The cause *forces* the effect to occur. Given a disposition and the right circumstances the disposition *cannot but* manifest, and the law *forces* or *necessitates* bodies to behave in some way rather than another. Furthermore, all three notions are closely tied to counterfactual conditionals. Laws support counterfactuals, causal relations can often be spelled out in terms of counterfactual claims, and the conditional analysis of disposition is at least a helpful device for the explication of dispositional properties. It thus seems to be reasonable to see how laws, dispositions, and causation are in fact related so as to give a unified account of these concepts and maybe of the kind of natural necessity involved.[4]

I will argue that at least most laws of nature are grounded in dispositions. Given a detailed understanding of these dispositions, we will also be able to understand how causal relations fit into a physical world.

2.1 Part–whole explanation

In this section I will explain why it is reasonable to assume that physical systems have dispositions and how to understand these dispositions.

I will argue that we need to assume that physical systems have dispositional properties because this assumption provides the best explanation for the way physics treats compound systems and their parts.

Let me start with an example of a part–whole explanation from quantum mechanics: carbon monoxide molecules consist of two atoms of mass m_1 and m_2 at a distance x. Besides vibrations along the x-axis, the atoms can perform rotations in three-dimensional space around its centre of mass. This provides the motivation for describing the molecule as a rotating oscillator, rather than as a simple harmonic oscillator. The compound's (the molecule's) behaviour is explained in terms of the behaviour of two subsystems, the oscillator and the rotator. In this case the parts are not spatial parts, they are sets of degrees of freedom. The physicist Arno Bohm, who discusses this example in his textbook on quantum mechanics, describes this procedure as follows:

> We shall therefore first study the rigid-rotator model by itself. This will provide us with a description of the CO states that are characterised by the quantum number n = 0, and will also approximately describe each set of states with a given vibrational quantum number n. Then we shall see how these two models [The harmonic oscillator has already been discussed in a previous

[4] For a discussion of these connections see (Handfield 2009: 4) and (McKitrick 2009: 31).

chapter. A.H.] are combined to form the vibrating rotator or the rotating vibrator (Bohm 1986: 128).

Thus, the first step consists in considering how each subsystem behaves if considered as an isolated system. The second step consists in combining the two systems.

This is a perfect illustration of a quantum-mechanical part–whole explanation. In carrying out this programme Bohm considers the following subsystems: (i) a rotator, which can be described by the Schrödinger equation with the Hamiltonian: $\mathbf{H}_{rot} = \mathbf{L}^2/2I$, where $\mathbf{L}$ is the angular momentum operator and I the moment of inertia; (ii) an oscillator, which can be described by the Schrödinger equation with the following Hamiltonian: $\mathbf{H}_{osc} = \mathbf{P}^2/2\mu + \mu\omega^2\mathbf{Q}^2/2$, where $\mathbf{P}$ is the momentum operator, Q the position operator, ω the frequency of the oscillating entity and μ the reduced mass.

He adds up the contributions of the subsystem by invoking a law of composition (for short: COMP):

IVa. Let one physical system be described by an algebra of operators, A_1, in the space R_1, and the other physical system by an algebra A_2 in R_2. The direct-product space $R_1 \otimes R_2$ is then the space of physical states of the physical combinations of these two systems, and its observables are operators in the direct-product space. The particular observables of the first system alone are given by $A_1 \otimes I$, and the observables of the second system alone are given by $I \otimes A_2$ (I = identity operator). (Bohm 1986: 147)

Thus, four laws are involved in this part–whole explanation:

(1) The law for the compound: The compound behaves according to the Schrödinger-equation with the Hamiltonian $\mathbf{H}_{\mathbf{comp}} = \mathbf{H}_{\mathbf{rot}} + \mathbf{H}_{\mathbf{osc}}$. (This is the *explanandum*; it describes the behaviour of the compound.)
(2) The law for the rotator: The rotator behaves according to the Schrödinger equation with the Hamiltonian: $\mathbf{H}_{rot} = \mathbf{L}^2/2I$.
(3) The law for the oscillator: The oscillator behaves according to the Schrödinger equation with the Hamiltonian: $\mathbf{H}_{osc} = \mathbf{P}^2/2\mu + \mu\omega^2\mathbf{Q}^2/2$.
(4) The law of composition (COMP) that tells us how to combine (2) and (3).

We explain the behaviour of the compound (the *explanandum*) as described in (1) in terms of (2), (3), and (4) (the *explanans*).

The solar system provides another example of a part–whole explanation. Even in the case of a highly integrated system as the solar system, you make the same steps. In the dynamic equations for the solar system (Lagrange-equation) you write the terms for the kinetic energy of the planets, etc. (that's the term that describes how the part would behave if it were on its own). You then combine these subsystems, i.e. you describe them by a single Hamilton function or operator. The only significant difference to the case above is that you need to add gravitational interaction terms. As in the case above, we rely on laws of composition, namely laws that tell us how to add up the various contributions.

Let me add as a final example of a compound system a falling stone in a medium. The subsystems are the freely falling stone on the one hand and the medium on the other. Both contribute to the overall behaviour.

2.2 An argument for dispositions

In this section I will argue that we need to assume that systems have dispositional properties in order to understand how part–whole explanations work.

As working definitions I employ the notions of categorical and dispositional properties as follows: A dispositional property is a property that, if instantiated by an object, is manifest under specific conditions only. A categorical property by contrast is a property that, if instantiated by an object, is manifest under all conditions. So, according to this distinction, categorical properties are limiting cases of dispositional properties. Note: In the limiting case of a categorical property, the distinction between a property and its manifestation doesn't do any work.

This is certainly not the orthodox way to draw the distinction, but the usual suspects fall on the right sides. Solubility and fragility, if instantiated by an object, become manifest under specific conditions. On the other hand, triangularity or massiveness (the best candidates for categorical properties), if instantiated by an object, are manifest under all conditions.

Why do we need dispositions to understand part–whole explanations? A basic ingredient of the explanans in a part–whole explanation is the reference to the behaviour the parts would manifest if they were on their own. However, this behaviour of the parts is not manifest while they are parts of a compound.

The vibrating rotator illustrates this: the subsystems contribute to the overall energy of the compound. But the behaviour of the subsystems is not manifest.[5] If they were manifest, the associated spectral lines could be measured—at least in principle. But they can't. The only spectral lines there are, are those of the compound.

The explanation relies on how the parts would behave if they were isolated, i.e. on how the rotator would behave if the oscillator were absent. The manifestation of the behaviour of the rotator is interfered with by the presence of the oscillator. The oscillator serves as an antidote with respect to the rotator's manifestation (and vice versa). It is, however, a *partial* antidote, because it does not suppress the manifestation completely. It prevents (complete) manifestation—however, it allows for *contributing* to the behaviour of the compound and thus partial manifestation.

The behaviour of the falling stone is—in general—not (completely) manifest either, because the medium serves as a (partial) antidote. Still, the behaviour of the falling stone in the absence of interfering factors plays an essential role in the explanation of the behaviour of the compound (falling stone in a medium).

[5] That is, they are not completely manifest. The distinction between complete and partial manifestation will be drawn in the next section. When I talk about manifestation simpliciter this has to be read as *complete* manifestation rather than partial manifestation (see sect. 2.4).

The description of the temporal evolution of the solar system relies among other things on how the parts (the planets and the sun) would behave if they were isolated (the kinetic energy terms). That behaviour is, however, not manifest. The other parts or planets as well as their interactions serve as partial antidotes.

To sum up: In part–whole explanations the behaviour of compound systems is explained in terms of the behaviour of the parts. The behaviour of these parts is not manifest because there are antidotes, namely the other parts or factors. The behaviour that we attribute to the parts is thus a behaviour that becomes manifest given certain circumstances obtain—the absence of antidotes or disturbing factors. A fortiori, we assume that the parts have dispositional properties. Their behaviour is not manifest under all circumstances (see Hüttemann 1998 and 2004. For a similar argument see Corry 2009).

2.3 Inertial and quasi-inertial laws

The above argument is not an argument for dispositional monism, i.e. the view that *all* properties are dispositional. I merely claim that at least some of the properties of physical systems, namely those appealed to in part–whole explanations, need to be construed as dispositional properties. Similarly, the argument does not claim that all laws need to be understood in terms of dispositions. According to the argument, those laws that describe the behaviour (properties) of systems that can play the role of parts need to be construed as attributing dispositions to physical systems.

The laws, which went into the explanations we looked at, are laws that describe the temporal evolution of systems: the temporal evolution of rotators, oscillators, stones, or planets. They describe *processes*. The temporal behaviour (or process) that such a law describes becomes manifest when there are no disturbing factors or antidotes, as for instance the medium or other systems. I will call these laws 'quasi-inertial laws' by analogy to Newton's first law:

> Every body continues in its state of rest or of uniform motion in a straight line, unless it is compelled to change that state by forces impressed upon it. (Newton 1999: 416)

The behaviour Newton's first law attributes to bodies becomes manifest, given nothing interferes. My claim is that many law-statements ought to be understood as making similar claims. Some temporal behaviour (process) is classified as in some sense 'inertial'. The law-statement says that unless something intervenes, some kind of behaviour will become manifest. So in a sense many laws, e.g. Galileo's law of free fall or the law that describes the behaviour of a rotator, can be classified as 'quasi-inertial' laws in the following sense: They describe what happens *if nothing intervenes*. They describe the *default*-behaviour of a system (or a default process).

2.4 Dispositions as contributors

The properties we rely on in part–whole explanations are dispositional in the sense outlined at the outset: The property/behaviour becomes manifest under special circumstances only. This is a feature they share (by definition) with dispositions we come across in everyday situations such as solubility or fragility. However, the dispositions appealed to in physical part–whole explanations have certain special features, which are not in general associated with everyday dispositions. These special features are responsible for the fact that my disposition-based theory of causation differs significantly from other dispositionalist accounts of causation (e.g. Molnar 2003 or Mumford 2009a).

Firstly, as already said, the behaviour appealed to in the part–whole explanations under consideration is typically temporal behaviour. The point is not that the manifestation itself takes time, but rather that the behaviour—when it is manifest—has a temporal dimension. It is a process.

Secondly, the triggering conditions for the dispositions appealed to are purely negative. The quasi-inertial behaviour becomes manifest in the *absence* of disturbing factors (antidotes).

Thirdly, the dispositions in question can be *partially* manifest; they *contribute* to the behaviour of the compound. Traditionally it is assumed the following possibilities obtain: either (i) a system S has a disposition D and D is manifest; or (ii) S has D, but D fails to be manifest; or, finally, (iii) S does not have D. In the case of various factors or parts contributing to a compound, a further situation has to be considered: S has D and D is partially manifest, i.e. it is not completely manifest but it contributes.[6] Partial manifestation and contribution may appear at first sight to be rather vague terms. This appearance is, however, deceptive in the case of physical part–whole explanation. It is in virtue of the laws of composition that contribution/partial manifestation can be made quantitatively precise. In our examples it can be made quantitatively precise how the rotator and the oscillator contribute to the compound system; similarly it can be made quantitatively precise how the medium affects a falling stone. So what I mean by calling dispositions 'contributors' is that the presence of a subsystem's disposition makes a difference to the behaviour of the compound and that the difference it makes depends on the law of composition.[7]

Let me add that my aim is not to provide a general theory of dispositions but rather to characterize those dispositions that we have to postulate in order to understand part–whole explanations in physics. The characteristics I have just highlighted are thus not to be taken to be those of dispositions in general.

[6] Why is partial manifestation not just a separate set of dispositions that are manifest in cases of non-isolation? Answer: That would not explain why it is the *same* disposition that is measured in different contexts (isolated and non-isolated cases alike).

[7] The notion of contribution developed here has to be distinguished from that of Molnar. According to Molnar, 'A manifestation is typically a contribution to an effect, an effect is typically a combination of contributory manifestations' (Molnar 2003: 195). Molnar's 'effect' is what I call 'the behaviour of the compound'; in my terms, the 'effect' is the manifestation of the dispositions of the compound. As I understand it, the contributions/manifestations for Molnar are real entities that mediate between the disposition and the overt behaviour (effect). I try to get along without such mediating entities.

3 Dispositions as contributors are not causes

In the special sciences and in everyday contexts we apply causal terminology. *Prima facie* it is difficult to see how we could do without. This raises the problem of how to reconcile the apparent indispensability of causal talk with Mach's and Russell's claims to the effect that causal statements are incompatible with what physics has to say about the world.

In section 2 I have argued that physical part–whole *explanations* presuppose that the properties/behaviours of physical systems that can play the role of parts ought to be understood as dispositional properties. One way to go from here is to argue for the thesis that causes are dispositions (Handfield 2009; Molnar 2003; Mumford 2009a). The idea is that dispositions cause their manifestations (I take this to be the standard dispositionalist account). I will, however, argue that this is a blind alley. While I am not denying that everyday dispositions such as fragility can be conceived of as causes of their manifestations, this move is not possible with regard to the dispositions introduced in section 2.

So the claim I am arguing for in this section is: Contribution and causation are different relations and thus not to be identified. Why is this so?

There may be cases of simultaneous causation, but the standard causal relation is a *temporal* determination relation: An *earlier* event causes a *later* event. Here a *temporal* determination relation is one in which a property, a state, or an event at t determines another property, state, or event at t^*, with $t^*>t$ (or $t^*<t$, though I will not consider this case). Given this definition, the part–whole relation and *a fortiori* the relation of contribution is an atemporal relation. The reason is that the laws of composition do not involve time.

As a consequence, causation (in general) cannot be identified with (reduced to) contribution (i.e. to the relation of a disposition to its manifestation). At best, *simultaneous* causation (if there is such a thing) can be identified with contribution. So we would have a theory for simultaneous (or instantaneous) causation and would need a different theory for cases in which the cause precedes the effect.

But maybe we should conversely consider contribution as a particular kind of (instantaneous) causation? The problem here seems to be that the alleged causes and effects are not in an appropriate way independent. The part–whole relation is constitutive. The relevant events, for instance, the rotator contributing to the compound and the compound being in a certain energy state, fail to be *distinct* events. It is, however, generally accepted that causes and effects have to be distinct events (cf. Lewis 1973: 165).[8]

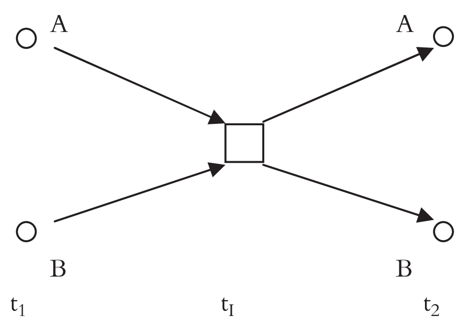
[8] Some authors such as Martin (2008), Molnar (2003), and Munford (2009a) argue that causation does not necessarily obtain between distinct events, but is rather a relation between identical events. 'It is not a matter of two events, but of one and the same event—a reciprocal dispositional partnering as a mutual manifesting. This surprising identity of what we had dimly thought of as the two-event cause and effect loses its surprise in the clear light of day' (Martin 2008: 46). There are several difficulties connected with this view. For our purposes the essential point is that we set out to explain our use of the causal terminology in the special sciences and in everyday contexts. This clearly involves that in general, the cause precedes the effect

So I conclude that this approach is a blind alley—dispositions (in our sense) do not cause their manifestations. It should be stressed that my argument relies essentially on the fact that the manifestation of the dispositions I consider is regulated by laws of composition, which do not contain a time parameter. It is for this reason that the standard dispositionalist account of causation is not tenable in a world as described by physics.

4 A disposition-based account of causation

We have just seen one attempt to ground the modal force of causes. The idea was to identify it with the disposition bringing about its manifestation. We have seen that this suggestion does not work. The dispositions in physics are contributors rather than causes.

More traditionally the modal force of causes (that in virtue of which the cause determines the effect to occur) has been grounded in laws of nature, conceived of as strict regularities. There are *laws* according to which the cause is a sufficient condition for the occurrence of the effect. As an illustration consider the following simple idealized case: Two billiard balls bounce against each other and are reflected.



Billiard ball A bouncing into B is the cause for B's deflection. According to the regularity view, the effect (B's deflection) had to occur in virtue of a regularity according to which the cause is a sufficient condition for the effect to occur. The law or regularity might be the following:

If at time t_1 billiard ball A is at x_{A1} and has velocity v_{A1} and B is at x_{B1} and has v_{B1}, then B at t_2 is at x_{B2} and has v_{B2}.

The problem with this kind of regularity account, as Russell has pointed out, is that there are not enough strict regularities. It is always possible that some factor interferes

and it is not easy to see how what we called temporal priority in section 1 emerges from Martin's and related accounts. Following Martin et al. would require us to reject a substantial part of the folk conception of causation—a rejection that will turn out not to be necessary.

(and as a matter of fact, such things do often happen). As a consequence, the alleged cause is not a sufficient condition of the occurrence of the alleged effect.

A second major account of causation is the counterfactual account. It explains how the cause partially determines the effect ('partial' because it allows for interfering factors): According to the counterfactual account, the cause partially determines the effect in the sense that if the cause had not occurred, the effect would not have occurred either. So with respect to the above example, the modal force of the cause is spelled out as follows: If at t_1 billiard ball A had not been at x_{A1} with velocity v_{A1}, B at t_2 would not have been at x_{B2} with v_{B2}. However, there are well-known problems with the counterfactual account of causation in general and with the account of modal force in particular. There are cases of causation and *a fortiori* of causes determining the effects to occur, in which the counterfactual conditionals do not hold (pre-emption).[9]

The challenge is thus to give an account of the modal force of causes that neither runs into the problem of interferences nor into the problem of pre-emption.

4.1 Central idea

The overall aim is to explain why in everyday life and in the special sciences we can successfully employ causal terminology despite Mach's and Russell's observations, which seemed to imply that the concept of causation is incompatible with fundamental physics. As I already mentioned, in this chapter I will focus on the problem of modal force. My account of the modal force of causation will proceed as follows. In a first step I will explain how causation and its modal force can be integrated into an idealized world as described by physics. In a second step I will explain how this accounts for our ordinary applications of causal terminology outside idealized physical models. The essential point with respect to the second step is that pragmatic considerations will play an important role. Finally, I will discuss some advantages of this view and relate it to other accounts of causation.

Let me start with a remark by Mach. 'In general we only feel the need to ask for a cause, if a (unexpected) change has occurred' (Mach 1896: 432). Similarly, Hart and Honoré have observed:

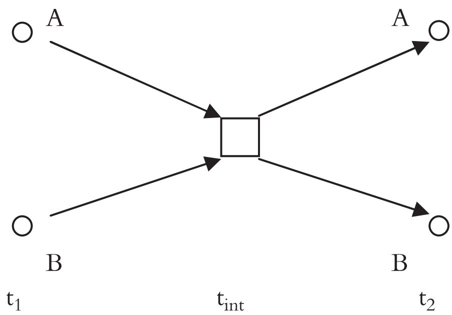
> The notion, that a cause is essentially something which interferes with or intervenes in the course of events which would normally take place, is central to the common-sense concept of cause. (Hart and Honoré 1959: 27)

Apparently, very often what we mean by applying the word 'cause' is that a certain factor disturbs a *default* process of some kind.[10] This observation will be my starting point and I will argue that it suffices to give an account of the modal force of causes that is at the same time compatible with physics.

[9] See Collins et al. (2004) for a discussion of these problems.

[10] Maudlin (2004) and Menzies (2007) develop similar ideas. Menzies has also developed an account of the kinds of context-sensitivities I discuss in section 5.

Let me return to the simple idealized case: Two billiard balls bounce against each other and are reflected.



We take the presence of the ball A at a particular place at t_{int} as the cause for B's deflection.

If we disregard the problem of asymmetry, we can explain in terms of physics why A's collision with B at t_{Int} is the cause for B's deflection.

B behaves according to Newton's first law, i.e. it has the disposition to simply continue in a straight line with uniform motion unless there is a disturbing factor. Newton's first law describes the inertial or default behaviour of B.

If the default behaviour does not occur, the law tells us that there must be some factor that interacted with B. The event of this factor interacting with the default process is the cause of the later non-occurrence of the default behaviour. Thus, in the idealized case we are considering here, the cause is something that prevents some system's default behaviour to occur.

> **In the idealized case under consideration, a cause is an actual disturbing factor (antidote) to the default behaviour that a system is disposed to display.**

(I will discuss in sections 5 and 6 how this definition has to be augmented in less idealized cases.)

As we have seen, laws are typically quasi-inertial or default laws. They tell us what happens if nothing interferes. Deflections, i.e. deviations from quasi-inertial or default behaviour, require an interfering factor—an interaction to have occurred. That is what the quasi-inertial laws of physics tell us. Causes are precisely these interfering factors that the quasi-inertial laws require in cases of deviations. Causes are thus the antidotes that explain why the dispositions that are ascribed by the quasi-inertial laws fail to be (completely) manifest.

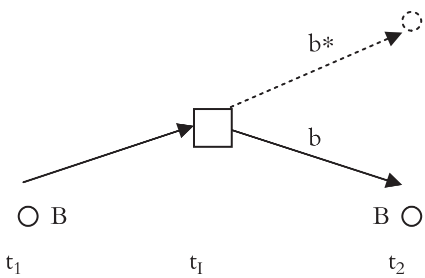
4.2 The origin of the causal counterfactuals

We can now explain why typically—though not always—counterfactual conditionals obtain according to which if the cause had not occurred, the effect would have occurred neither.

$\sim O(c) \Box\!\!\rightarrow \sim O(e)$ (Lewis (1986b): 167).

Why is this counterfactual true in many cases of causation? Consider again our simple example:

My claim is that we hold this counterfactual true in virtue of Newton's first law: If A had not collided with B, B would have taken path b* rather than path b—as it actually has (see Figure 2).



b: B's actual path
b*: the path B would have taken if A had not interfered at time t_I

Laws of nature describe how systems would behave in the absence of disturbing factors, i.e. they attribute dispositions to physical systems. Laws of nature usually describe counterfactual situations and should therefore be read as saying, for instance, 'If the hydrogen atom were isolated it would behave according to the Schrödinger equation with a Coulomb potential.' According to my proposal, it is exactly the underlying dispositions that make true the relevant counterfactuals. So we can understand why in many cases of causation, causal counterfactuals are true if we assume that the systems in questions have the relevant dispositions (see Hüttemann 2004: 110–15).

4.3 Pre-emption and interference

I promised an account of causation that does not run into the problems of interference and pre-emption. So how are these problems evaded?

As is well known, sometimes backup causes may be around. We do think that backup causes do not undermine causation. However, they undermine counterfactual dependence. Since according to our account causation is not identified with counterfactual dependence, there is no problem of (either early or late) pre-emption. Counterfactual dependence is not necessary for causation. For this solution of the pre-emption problem it is essential that the cause, i.e. the occurrence of the disturbing factor or antidote, can be spelled out in terms of *actual* facts about interactions rather than in terms of counterfactual claims about what would have happened if the factor in question had been absent. All that is needed for singular causation is the disposition of a system to behave according to a (quasi-)inertial law and the *actual* disturbing factor. The presence of *poten-*

tial disturbing factors is irrelevant on the account presented here (see Maudlin (2004) for a similar approach).

How does our account evade the interference problem? After all, laws play quite a significant role in our account. The main move is to understand laws as ascriptions of dispositions rather than as strict regularities concerning manifest behaviour. Newton's first law is not a strict regularity concerning manifest behaviour. It is no strict regularity because in fact there are often (maybe always) interfering factors around (impressed forces). However, strict regularities are not needed for our account. All that we need for our account of causation are laws based on dispositions that tell us what would happen if nothing disturbs the default behaviour. The problem of interferences does not apply.

5 Causal fields

The example we have considered so far was highly idealized. Relative to this idealized setting a cause was defined as a disturbing factor to the default behaviour that a system is disposed to display. The situation was idealized because we have abstracted away from diverse factors. In non-idealized situations there are always further interference factors besides what we identify as 'the cause', e.g. the molecules in the air collide with the billiard ball in question. Furthermore, certain constitutive conditions obtain, such as the presence of the billiard-ball table.

John Mackie has described this as the *causal field*:

> Both cause and effect are seen as differences within a field; anything that is part of the assumed (but commonly unstated) description of the field itself will, then, be automatically ruled out as a candidate for the role of cause. (Mackie 1980: 35)

It is for instance taken for granted that the gravitational field is stable, the history of the universe is kept fixed, etc. It is only relative to these and other background assumptions that a default process can be defined. If, for example, the billiard-ball table had not been smooth but bumpy, and if this were part of the background assumptions (causal field), the default process would have been different.

It is important to notice that in one and the same situation different people may make different assumptions about the background. In a case of a car accident one observer might take the dirt on the street as part of the background. The default process in this case might be that the car starts to skid slightly and that the driver is able to stabilize the car afterwards. The cause for the accident is the driver's drunkenness and his/her inability to control the car in difficult circumstances. Someone else might take the drunkenness of the driver as part of the background conditions. The default process is the driver driving the car home safely (provided no difficult situations occur). In this case the dirt on the street is the cause of the accident.

What counts as a quasi-inertial or default process and what counts as a cause will be determined relative to a causal field and thus relative to pragmatic or subjective concerns. However, this relativity does not imply that quasi-inertial processes are arbitrary.

If one has chosen a particular field, whether or not a certain process is an inertial process is an objective matter. If the smoothness of the billiard-ball table is part of the causal field, it is no longer a subjective matter whether the default process of the billiard ball consists in a uniform and rectilinear motion rather than in some kind of non-uniform motion.

To sum up: An augmented definition of 'cause' has to take the causal field into account:

> **Relative to a causal field, a cause is an actual disturbing factor (antidote) to the default behaviour that a system is disposed to display.**

The introduction of causal fields allows us to explain some of the features that are usually associated with the causal relation.

First, it allows to distinguish between causes and conditions. The smoothness of the billiard-ball table (as part of the causal field) is a mere condition of the deviation of ball B, whereas the interaction with A is its cause.

Second, it allows causes to be local. While it may be true that the conjunction of all relevant factors for a certain event is non-local in the sense that an indefinite environment has to be taken into account, the introduction of the causal field allows to relegate the environment to the field. Thus, the intuition that causes are local can be respected.

It is, however, a consequence of this conception that whether or not a particular event is considered to be part of the causal field—and thus a condition rather than a cause—is a matter of pragmatic and subjective concerns. Whether or not an event is a cause is thus not an entirely objective fact. This consequence seems, however, to capture how we actually employ causal terminology (as illustrated in the case of the car accident). Objectivity seems to be an issue where the folk-conception and the actual use of the terminology point in different directions.

6 Limiting conditions

What we have shown is that—given our conception of causation—causal terminology is applicable in certain cases (such as the billiard ball example). However, we have no guarantee that the terminology is generally applicable.

I would like to suggest that causal terminology is applicable given that certain limiting conditions obtain, but that it might not be applicable in other cases.[11] What are these limiting conditions?

In some compound systems (such as the system consisting of the two interacting billiard balls) we can identify two subsystems and their default behaviour (because they are isolated prior to their interaction and afterwards), as well as their spatio-temporally localized interaction. If these conditions are met, the application of the causal terminol-

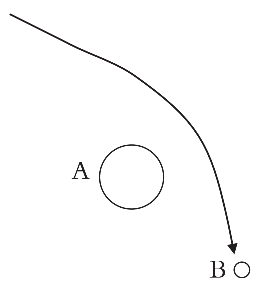
[11] Dennis Dieks has developed similar ideas in his dissertation (Dieks 1981, chap. 3). For a recent articulation of this strategy see Norton 2007.

ogy is plausible. In other words: As long as the (compound) system we are investigating can reasonably be described as having the 'two colliding billiard balls'-structure (localized interaction, isolated systems prior to and after the interaction), causal terminology is applicable.

While the above-mentioned conditions are sufficient for the application of causal terminology, they do not seem to be necessary. For instance, in real cases, to which we apply causal terminology, the disturbed systems are never completely isolated. The billiard balls collide with molecules in the air, the moon exerts a gravitational pull, etc. Some of these factors can be relegated to the causal field while others are simply neglected. Thus, classifying certain interfering factors as negligible is a further pragmatic aspect that is presupposed by the application of causal terminology.

Furthermore, the interaction we considered in the billiard-ball case was local (collision). That seems not to be a necessary condition for the applicability of causal terminology. Consider the following case:

A compound system consisting of a large massive system A and small massive system B that is deflected:



Let us assume that A simply sits there and by gravitational attraction deflects the path of B. Even though the interaction is not local, we would still say that the gravitational potential generated by A is the cause of B's path being deflected. The interaction is always on and thus B never actually instantiates the default or quasi-inertial behaviour. Still, since in this situation B approaches inertial behaviour if it is very distant from A, we can understand why in this case causal terminology is applicable. In this case the essential point seems to be that we have theoretical and/or experimental means on the basis of which we can attribute a certain default behaviour to B that is then disturbed by A.

In the end there may not be a clear-cut line that distinguishes cases in which the terms 'cause' and 'effect' are applicable from those in which they are not. There are, however, clear cases. We have come across systems to which the causal terminology is clearly applicable. On the other hand, there are, for example, closed deterministic systems with nothing remotely similar to the billiard-ball structure. They seem to be clear candidates where the application of causal terminology is inappropriate. It was these cases that Russell had in mind:

> The law of gravitation will illustrate what occurs in any advanced science. In the motions of mutually gravitating bodies, there is nothing that can be called an effect; there is merely a formula. Certain differential equations can be found, which hold at every instant for every particle of the system, and which, given the configurations and the velocities at one instant, or the configurations at two instants, render the configuration at any other earlier or later instant calculable. That is to say, the configuration at any instant is a function of that instant and the configurations at two given instants. […]. But there is nothing that could be properly called 'cause' and nothing that could properly be called 'effect' in such a system. (Russell 1912–13: 14)

Russell's claim was of course meant not only to apply to this special case. However, as we have seen, there are limiting conditions such that the application of causal terminology is plausible. The limiting condition we have identified is this: It must be plausible to attribute default behaviour to a subsystem of a compound and to identify a disturbing factor for the deviation of the default behaviour. And this does indeed often seem to be case, at least relative to certain pragmatic considerations.

7 Relation to other process-theories

I argued that the application of causal terminology can be best understood if a cause is taken to be an actual disturbing factor (antidote) to the default behaviour that a system is disposed to display (relative to a causal field). This is a claim about processes, because the default behaviour concerns the temporal evolution of a system (e.g. the process that is described in Newton's first law). So how exactly is the disposition-based view related to traditional process-theories?

Process-theories tend to take 'causation to be the transfer or persistence of properties of a specific sort' (Dowe 2009: 214). The default processes we have talked about can indeed be characterized in terms of the persistence of properties: The behaviour in question is persistently manifest as long as nothing intervenes. Process-theories consider interferences with these processes as cases of causation. Thus far I agree. There are two important disagreements: First, I do not take the processes themselves to constitute causation. A statue being at a certain place at 2 p.m. today is not a cause of its being there at 5 p.m.—even though it may be a condition. Therefore, I do not talk about *causal* processes but rather about *default* processes. Second, the characterization of the relevant processes and disturbances (interactions) is different. Whereas Dowe and Salmon characterize these processes either by the mark-criterion or in terms of invariant or conserved quantities (Dowe 2009 provides an overview), I characterize them in terms of the underlying systems' dispositions. A ball rolling on a flat surface is classified by traditional process-theorists as a causal process because it conserves kinetic energy and momentum. I characterize it as a default process, because it manifests a disposition, namely the one that is attributed in Newton's first law. Having certain invariant/conserved physical properties is thus not necessary to qualify as a default process. The essential question is whether or not the relevant system has a certain disposition. It is for the physicists to decide whether or not bodies have the disposition to continue in uniform rectilinear motion if no forces are impressed. Similarly it is for economists to decide whether or not (within a certain causal field) an economic system in which inflation rises will yield higher unemployment rates (if nothing interferes).

As a consequence whether or not something qualifies as a default process need not be spelled out in terms of physics. To the extent that biology or economics attributes dispositions to systems that concern their temporal evolution, these disciplines are dealing with biological or economical default processes.

Similarly, what qualifies as a disturbance of a default process needs to be spelled out in terms of physics. I want to leave room for various kinds of disturbances that have to be specified locally. It is the job of the sciences in question to provide a more detailed description of the disturbance. While physics considers a certain repertoire of disturbance factors that can be described in terms of physical interaction or, maybe, conserved quantities, economy considers state interventions, decisions of the Federal Reserve Bank, or natural catastrophes.

The disposition-based process-theory thus promises to take account of the fact that causal relations may obtain on various 'levels'.

A further advantage of the disposition-based process-theory vis-à-vis traditional process-theories is its ability to cope with double prevention cases. Ned Hall has described a paradigm case:

> Suzy is piloting a bomber on a mission to blow up an enemy target, and Billy is piloting a fighter as her lone escort. Along comes an enemy fighter plane, piloted by Enemy. Sharp-eyed Billy spots Enemy, zooms in, pulls the trigger, and Enemy's plane goes down in flames. Suzy's mission is undisturbed, and the bombing takes place as planned. (Hall 2004: 241)

We want to say that Billy's pulling the trigger is a cause of the bombing of the target. Traditional process-theories have a problem with this example because there is no continuous physical process leading from Billy's pulling the trigger to the actual bombing.

Here is how the situation can be described given the disposition-based process-theory:

Let's start with the simple case. Suzy's bombing being the cause for the destruction of the target. Here the *target sitting around peacefully* is the default process that is disturbed. Second case: If Enemy would prevent Suzy from bombing the target, we consider *Suzy bombing the target* as the default process that is disturbed by Enemy's intervention. So, Enemy's intervention is the cause of the target's survival. Third, if Billy shoots down Enemy, the default process we are considering is *Enemy preventing Suzy from bombing the target*. This process is disturbed by Billy's pulling the trigger. Billy's pulling the trigger is therefore the cause that *Enemy preventing Suzy from bombing the target* does not take place and thus a cause of the bombing of the target.

So the disposition-based process-theory is able to cope with at least some objections to traditional process-theories.

8 Probabilistic causation

One might wonder how the account outlined so far is relevant for causal claims in the special sciences, since many of these are probabilistic causal claims. Probabilistic causal claims such as 'Smoking causes lung-cancer' are typically type-level causal claims in contrast to the token-level claims I have considered so far. Thus the question is whether and how probabilistic causal type-level claims fit into the picture outlined so far.

There are two different ways in which the truth of such claims may depend on underlying processes and disturbances, depending on whether the underlying processes are deterministic or genuinely indeterministic.

(1) Deterministic case: The probabilistic type-level claim obtains in virtue of coarse graining over different kinds of different causings. Suppose the claim is that some event type X probabilistically causes Y to occur. This may be true in virtue of the fact that two different kinds of situations are involved:

situation type A: X disturbs process P, Y occurs as a result of ensuing process P*.
situation type B: X disturbs process P, X* disturbs ensuing process P*, Y does not occur.

Coarse graining over heterogeneous situations of type A and type B makes probabilistic type-level claims true. If $Pr(Y|X) > Pr(Y)$ one might say that X is likely to cause Y. In this deterministic case no new element has to be introduced in order to understand probabilistic causal claims at the type-level.

The situation is different if one wants to accept genuine chancy causation:

(2) Indeterministic case: The probabilistic type-level claim obtains in virtue of genuine chancy causation. Suppose the chance of receiving a certain illness I is 0.5 per cent. Excessive consumption of X raises the probability to 2 per cent. We furthermore assume that the relevant type-level claim 'Excessive consumption of X *causes* I' cannot be explained in terms of coarse-graining over heterogeneous situations. This case requires the introduction of genuinely indeterministic processes, i.e. processes that have various outcomes and a probability distribution over the outcomes. It might be the case that human beings (in the usual contexts) develop illness I in 0.5 per cent of the cases and non-I (i.e. fail to develop I) in 99.5 per cent of the cases. Take this to be the default process. I take it that x is the cause of i (lower case for instantiations of types) if the person in question excessively consumes x, if the person furthermore develops I, and if the original probability distribution changes in virtue of the consumption of x such that it is more likely to develop I.

Genuine chancy causation (at the token level) can then be understood as follows:

x causes i, if x occurs, i occurs, and the default process probability distribution has been disturbed by x so that $Pr(I/X)>Pr(I)$ is true.

The central idea is that genuinely indeterministic processes have a probability distribution over outcomes, which might be disturbed by interfering factors. If the disturbance raises the probability of the outcome, then the interfering factor may be said to have caused a certain outcome. This is an extension of the original account to the case of chancy causation. There are certain well-known difficulties with the probability raising requirement (see e.g. Hitchcock 2011); however, I hope the sketch suffices to indicate how genuine chancy causation may be adumbrated.

Probabilistic causal type-level claims in the special sciences may thus be true in virtue of underlying disturbances of deterministic processes or in virtue of disturbances of indeterministic processes or both.

9 The origin of causal modality

In this chapter I have argued that causation can be explained in terms of laws that describe default processes and that these laws should be understood in terms of dispositions. What does that tell us about the origin of causal modality or causal determination?

In the case of the billiard balls the origin of this determination becomes apparent if we consider the situation as the behaviour of a compound system. What is relevant in this case is the law of the conservation of momentum. A's state and the law of the conservation of momentum determine how B behaves.

More generally: The law for the compound system (plus the states of the other parts) determine the effect, i.e. the behaviour of the part that is disturbed.

The modal force or determination is thus due to those laws that determine the behaviour of the compound system.

(N) The effect occurs necessarily because the compound system manifests its behaviour necessarily.

But what kind of necessity is involved here?

Let me present a somewhat speculative sketch that provides an answer to this question. In what follows I take metaphysical necessity to be necessity that obtains in virtue of the essence of a system. So if a system has a certain property with metaphysical necessity it has it in virtue of its essence. Now let us assume that the systems we are considering have their dispositions essentially. This will not yet give us the explanation of the modal force or determination we are looking for (N).

The essential point is that two issues have to be distinguished:

a) the claim that an object or property has a disposition necessarily (dispositional essentialism) and
b) the claim that a disposition displays its manifestation necessarily.

Dispositional essentialism (a) does not explain why the dispositions in question manifest their behaviour necessarily. In fact, they don't. As long as antidotes are possible, manifestation cannot occur with necessity, i.e. it is still not clear what kind of necessity is involved in (N) (for a discussion of these points see Schrenk 2010).

A possible solution to this problem is the following: What happens if the compound (consisting of the disturbed system and the interferer) is itself disturbed, is determined by the disposition(s) of this compound, the disposition(s) of the new interference factor as well as the law(s) of composition. Now assume that it is part of the *essence* of systems that they behave according to the laws of composition. Then the following holds: It is necessarily the case that, if nothing interferes, the compound system manifests its behaviour. It is furthermore necessarily the case that, if something interferes, the system displays a different behaviour of this or that kind. In other words, if we assume that the laws of composition hold with metaphysical necessity, then the manifestation of the dispositions of the systems happens with conditional metaphysical necessity. The conditions in question are the disturbing factors. So it holds with metaphysical necessity that—if nothing interferes—billiard ball A deflects billiard ball B. The modal force of causes obtains in virtue of the essence of the systems involved. The necessity involved in (N) is conditional metaphysical necessity.

The assumption of conditional metaphysical necessity allows us to give a unified account of dispositional, nomological, and causal modalities. It is conditional metaphysical necessity that is at work in all of these cases. If the systems have not only their dispositions essentially but furthermore conform to the laws of composition in virtue of their essence, then, first, with metaphysical necessity—if nothing interferes—the disposition becomes manifest. Second, with metaphysical necessity—if nothing interferes—the law that attributes the disposition in question is true. And, third, with metaphysical necessity—if nothing interferes—the disposition of compound systems such as the system consisting of the two billiard balls becomes manifest, i.e. with metaphysical necessity—given the right conditions—the cause determines the effect.

10 Conclusion

To sum up: I have asked in this chapter how and under what circumstances causal terminology can be applied to the physical world. The focus was the question how to understand that the cause determines the effect to occur (I disregarded the problem of the asymmetry of causation). I argued that we should conceive a cause as an actual disturbing factor (antidote) to the default process that a system is disposed to display (relative to a causal field). Given this disposition-based process-theory we can understand how causation is part of the physical world. We can furthermore understand how causation can have many of the features that the folk-conception of causation associates with it: modal force, locality, the distinction between causes and conditions, and the multi-level character of causation. The idea that whether or not something is a cause (as opposed to a condition) is an objective matter could not be vindicated. It should also be stressed, that causal terminology is only applicable given certain limiting conditions that need to be realized by all kind of (physical) systems. [12]

[12] I would like to thank two anonymous referees, the audience in Nottingham, the members of the DFG-funded Research Group FOR 1063 on *Causation and Explanation*, as well as John Roberts for their very helpful comments on an earlier version of this chapter.