Causal Structures in Quantum Field Theory

Jakob Sprickerhof

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

School of Philosophy, Religion and History of Science

September 2015

Acknowledgments

I wish to thank all those who gave me the opportunity to discuss my work with them and who have helped in the preparation of this dissertation, in particular: Michael Esfeld, James Fraser, Klaus Fredenhagen, Michael Heidelberger, Holger Lyre, Matthias Neuber, Alexander Reutlinger and Juha Saatsi. Above all, I wish to thank Steven French for his continuous support and countless invaluable comments.

J. S.

Abstract

In this dissertation, I will argue that quantum field theory can be interpreted as a description of causal interactions. I will first discuss previous theories of physical causation, and come to the conclusion that they are inapplicable in quantum field theory. As a consequence, I will start to develop a new theory of causation by first analysing the concept 'causation' into the most basic and widely shared intuitions, in order to find out later whether these intuitions can be reduced to physics. I will then have a closer look on the intuition that causation is a directed relation, which is commonly regarded as incompatible with the symmetries of physics. I will present a new argument to the effect that causation and quantum field theory are compatible with respect to the directionality of causation. After that, I will analyse the theoretical description of interactions in quantum field theory, and in particular group structure, locality and local conservation laws will crystallise as the central concepts that a causal interpretation might be based on. Subsequently, I will discuss and present replies to what I believe are the most relevant objections to a causal interpretation of physics, namely, Haag's theorem, the measurement problem and entanglement. In the final chapter of this dissertation, I will conjoin the results of the previous chapters to a new theory of causation for quantum field theory. The main result will be that a causal process is a quantum field theoretical interaction, i.e., the exchange of energy from an initial to a final state via a force.

Contents

Acronyms

AQFT: Algebraic quantum field theory

C: Charge conjugation

CD: Causal directionality

CI: Causal intersection

CND: Causation and nomic dependence

CQ1, 2: Conserved quantity 1, 2

CQT: Conserved quantity theory of causation

$\mathbf{CQT_{QFT}}$: Conserved quantity theory of causation for quantum field theory

EDR: Rigid existential dependence

EPR: Einstein-Podolsky-Rosen

GRW: Ghirardi-Rimini-Weber

IC: Immanent causation

IN: Intervention

IV: Intervention variable

LSZ: Lehmann-Symanzik-Zimmermann

M: Manipulability theory

MT: Mark transmission

NC: Necessary condition

NST: No-signalling theorem

P: Parity

PCI: Propagation of causal influence

PND: Physics and nomic dependence

QFT: Quantum field theory

QM: Quantum mechanics

SC: Sufficient condition

STR: Special theory of relativity

T: Time reversal transformation

TC: Total cause

TD: Temporal directionality

¬TD: No temporal directionality

1 Introduction

Nature will always maintain her rights, and prevail in the end over any abstract reasoning whatsoever.
(Hume, An enquiry concerning human understanding. V, §2)

'The United Nations adopted the Universal Declaration of Human Rights in 1948, *because* of the experience of the Second World War. Some animals have developed echolocation through evolution, *because* it is an effective means to communicate and hunt. The tides come and go, *because* of the moon revolving around the earth. The sky is blue, *because* photons of the sunlight of different wavelength scatter differently on the molecules in the atmosphere.' – All of these statements are similar in that they cite a reason for an event or phenomenon, and arguably not just any reason, but a cause. It is obvious that there is a continuity, from history, over biology to cosmic physical processes and finally to sub-atomic physical processes; the statements might be different in topic, but that is all – right? Of course, it is not that simple. The discussion among philosophers has placed serious doubt on this continuity. Especially in recent decades, it has become commonplace to deny that physical theories describe causes and their effects. Even among proponents of causation, the standard view is that there is something special about physics that makes this science hostile to causal processes. Is this true? Are fundamental physical processes really that different from processes that are closer to our everyday experiences? Do we need to think of fundamental physical processes in a radically different way compared to affairs in everyday life, without causes and effects? This, by and large, is the question that concerns me in this dissertation, though, I will narrow it down in a minute.

Dissertations rarely start from zero; some assumptions have to be presupposed. There is a long tradition in philosophy that is now known under the names of naturalism or naturalised metaphysics (cf. Kitcher 1992). This tradition can hardly be summarised into one or very few positive claims, however, it is united in its rejection of an Aristotelian prima philosophia. It is the epistemological claim that systematic knowledge about the world is only produced by the natural sciences, and not philosophy. The role of philosophical investigation is then not even that of an ancilla scienciae, to variate a well-known aphorism, but it can merely rephrase, analyse or systematise scientific knowledge, without producing any genuinely new knowledge. This, of course, is only a brief statement of the core of naturalism, and whether this position is justifiable and coherent I cannot here argue for, but have to presuppose. I will adopt a naturalistic stance throughout this dissertation.

Within this tradition, it is feasible that philosophical enquiry should be based not simply on any scientific theory, but on the best and most advanced scientific theory available. For example, an ontology of Newtonian mechanics might not be entirely false, but is not as good as an ontology of quantum physics. The picture of science, underlying this claim, is that of an accumulative progression, moving towards the truth. Again, I assume that there is a sufficient consensus about what that means to liberate me from the burden of any further explication. Embedded is this picture in the position of scientific realism, according to which science gives approximate true knowledge of the world. Throughout this dissertation I presuppose that scientific realism is true.

Among philosophers, who largely share the above assumptions, it is now commonplace to accuse others, often belonging to the same group, of not paying sufficient attention to our most advanced sciences. I share this critique, and think that at least within the topic that I am concerned with, many authors either openly or tacitly are not sufficiently connected to our best theories. It is my aim to take part in bridging this gap between philosophy and today's natural sciences.

In the above paragraphs I have outlined the broad background of this work. On top of that, every work like the present one needs to narrow down its topic to a manageable level. The first restriction I have to impose is with respect to the sciences. Arguably, there is a variety of sciences that produce knowledge of the world: physics, biology, neurology, chemistry and so on. Of these, I will only be concerned with physics. The reason for this choice is the intuition (or hope) that every other science in a certain relevant sense can be reduced to physics. This is at present a highly contentious and speculative claim, for which I do not even dare to start arguing. Second, within physics there are many theories that might count as contenders for the current best theory, such as quantum field theory (in various formulations), general theory of relativity, string theory, loop quantum gravity and so on. Of these, I will only be concerned with quantum field theory (QFT) as it is found in the common textbooks that most advanced physics students have to study at one point. The reason for this choice is that QFT has had immense success over the last hundred years and is now the theory that is most established as the description of very small processes. The general theory of relativity, of course, shares these attributes for very large processes, which brings me to a general point. By focusing on QFT, I do not wish to dismiss any other well-established science. The project of the present work could just as well have been carried out in biology or the general theory of relativity. I have briefly given a reason for the restriction to QFT, but no argument. Hence, the focus on QFT might be regarded as a choice of personal preference that is constrained by the spatiotemporal restrictions of a dissertation.

The second general choice that necessarily has to be made is with respect to the philosophical topic. Among the many areas in philosophy on which the natural sciences have an impact, I will focus on the philosophy of causation, which again might be regarded as a choice of personal preference. Notwithstanding, I have reasons for why I regard causation as a topic worth spending some time with. For our understanding of causation, a guiding idea seems to be that causation is more than correlation. There is a difference in the content of the statements that, say, 'I wake up every morning at sunrise' and 'I wake up every morning, because of the sunrise'. If this is correct, then there are at least two reasons to care about causation. According to Hall (2011, p. 97), "the value of a concept of causation derives from details of our actual human predicament. First, we need *control* over our world. Second, we need to *understand* it." This means that first, causation is useful to delineate effective from ineffective strategies of control, as Nancy Cartwright has argued in numerous publications. For example, if I want to keep the tire of my bike from flattening all the time, then it probably does not help me to know that it usually flattens during daytime. Second, causation increases our understanding of the world, in that it answers a why question, viz., causes explain. There is a strong intuition that we understand a phenomenon better if we know why it happened, as compared to merely knowing which other phenomena happened at roughly the same time and place.

From these assumption and motivations the main question of the present work emerges:

Can QFT be understood as describing causal processes?

In order to find an answer, it will be necessary to first gain a better understanding of what we mean by 'causation', not least to simply be able to distinguish the causal form the non-causal. Second, QFT has to be analysed to single out bits and pieces that might be interpreted as causal. The latter can, of course, not be mistaken for actual research in QFT. Even though I might argue for an understanding of 'causation' that many will not share, everything I have to say on QFT can be found in textbooks, and, at least from a physicists perspective, should be uncontroversial.

According to naturalism, if it turns out that a natural science tracks causal relations, then having knowledge of causal relations is nothing extra to having knowledge of that natural science. This does not mean that all scientific knowledge is about causal relations. Furthermore, it does not mean that there is one theory of causation that summarises all possible causal relations. Even though in this work I will present one specific theory of causation, I prima facie do not oppose causal pluralism that can be plural on two axes: First, it might be the case that there are different concepts of causation in different sciences. Second, it might be the case that there are different concepts of causation in one single science (cf. Godfrey-Smith 2009; Hitchcock 2003; Ross & Spurrett 2007; Williamson 2006). Nothing I say precludes either form of pluralism or a combination thereof. I will, however, at some point argue against certain interpretations of 'causation' for different reasons.

In developing a theory of physical causation, one faces a battle on three fronts. First, causation by itself is a highly controversial concept. There are uncountable many different theories of causation and confusingly many arguments surrounding them. Second, even though there might be a relatively uncontroversial textbook formalism of QFT, there are different formulations of QFT that potentially lead to different interpretations. Furthermore there are many potentially problematic issues surrounding the textbook version. Third, there are special problems that a causal interpretation of physics faces, and it is today the predominant view among philosophers that both, causation and physics, are incompatible. I will address all three issues throughout this work, however, it is practically impossible to win the battle on all three sides; all I can do is to discuss what I regard as the most relevant problems.

I wish to end this introduction with a brief summary of each chapter:

- **Chapter 2:** I give an overview over all theories of physical causation that have been proposed so far. All of them have in common that they regard causation to be the physical transfer of some conserved quantity from cause to effect. An exeption, however, is Salmon's process theory, where the possibility of transmitting a mark is the distinct feature of causal processes. The best developed theory of physical causation will turn out to be Dowe's conserved quantity theory (CQT). I will analyse how far the CQT is compatible with QFT and come to the conclusion that it cannot be applied to QFT in its current form.

- **Chapter 3:** I argue that it is methodologically not feasible to conjoin the CQT and QFT simply by carrying out small adjustments to the former. Rather, a new theory of causation has to be developed from the very beginning. In order to do so, it is necessary to follow a clear method. I outline two methodologies, the Canberra plan and naturalism, and show that they propose the same method to the extent that is relevant here. Accordingly, the method I will follow comprises of two steps: First, the analysis of the concept 'causation' into clearer and widely shared intuitions. Second, the reduction of these intuitions onto physics. I carry out the first step by discussing in detail the most common intuitions on causation found in the literature. I conclude that causation is a relation between distinct events that can be exploited for manipulation, has a direction and is stable under interventions.

- **Chapter 4:** The most prominent reason for why causation and physics are commonly seen as incompatible is the conflict between the directionality of causation and the symmetry of physics. I show that there are two separate arguments involved: first, an argument based on the incompatibility between the temporal asymmetry of causation and the time reversal symmetry of physics, and second, and argument based on the incompatibility between causal asymmetry and the symmetry of determination in physics. I show that the first argument fails, and moreover, that there is a temporal asymmetry of the physical facts onto which the temporal asymmetry of causation can be reduced. With respect to the second argument, I show that it fails, if causal asymmetry is understood in terms of asymmetric dependencies between models of causal structures, and determinism is understood in a local, model relative sense. Thus, this chapter completes the second step of the naturalistic method with respect to the directionality of causation.

- **Chapter 5:** This chapter is concerned with the analysis of QFT, which lays the basis for the second methodological step for the remaining intuitions. I present three case studies in order to find out how interactions are described in QFT. Building up on this, I take a closer look at five characteristics of QFT that seem promising as a reduction basis for causation: group structure, local conservation laws, the principle of locality, force fields and probabilities. I show that, (i) following Wigner, group theory orders states in QFT into kinds; (ii) local conservation implies that conserved quantities cannot change their location from one region to another without a conserved current between them; (iii) locality implies that there is no action at a distance, (iv) a realist view on forces is justified, but not on virtual particles, (v) QFT is a probabilistic theory, in that interactions only happen with a certain probability.

- **Chapter 6:** I discuss the most threatening aspects of quantum physics for a causal interpretation: Haag's theorem, the measurement problem and entanglement. I argue that Haag's theorem does not apply to renormalised QFT and to Haag-Ruelle scattering theory, and therefore can be dismissed here. With respect to the measurement problem, I argue that the common position that it cannot be solved by physics, but only by philosophy is unwarranted. Consequently, I motivate a positive attitude according to which the measurement problem might still be solved by a quantum theory that does not deviate to a relevant extent from today's QFT. Thus, I acknowledge that the measurement problem is potentially dangerous, but also urge that more research in physics has to be done before any final conclusions can be drawn. Finally, I argue that the correlations in EPR-type experiments, which are due to entanglement, do not support any causal interpretation. In particular, non-separability makes it impossible do identify distinct causal relata, and the spacelike separation of the measurements prevents any possible relata from having a fixed temporal order, which is incompatible with the directionality of causation.

- **Chapter 7:** I complete the second methodological step by showing how the intuitions from chapter 3 can be reduced to the characteristics of QFT from chapter 5. In brief: (i) The causal relata are defined via group structure, and given in initial and final states. The causal relation is given by a force that transfers energy. (ii) The probability for a cause to bring about the effect is given by the S-matrix. (iii) We can manipulate the energy of a final state by manipulating the energy of the initial state. (iv) The directionality of causation, temporal and causal, is represented by local energy flux and dependencies in causal structures, as argued in chapter 4. Finally, I combine these results into a new theory of causation, $\mathrm{CQT}_{\mathrm{QFT}}$: A causal process is a quantum field theoretical interaction, i.e., the exchange of energy from an initial to a final state via a force.

2 Towards a theory of physical causation

2.1 Introduction

Since the 1970's a variety theories of causation have been proposed that intend to identify causation with a physical process or interaction. All of them make use of conserved quantities, either to delineate causal from non-causal process or to describe the causal relation. This chapter is an assessment of where this project stands today, of what has been established and what remains questionable. I will therefore introduce theories of physical causation and critically analyse their tenability. The emphasis here will be on the compatibility with quantum physics, and conceptual issues will only be discussed insofar the internal coherence of a theory is in question. However, I will merely point out what I believe are the most relevant problems and leave their full discussion for later. In short, this chapter will form the basis for the further investigation in the present work, by showing what has been done and which issues need to be discussed further.

The theories that I am going to present are the transference theories of Aronson (1971), Fair (1979) and Heathcote (1989), the mark transmission theory of Salmon (1994) and the conserved quantity theory (CQT) of Dowe (2000). The emphasis will be on Dowe's CQT, as the most recent and elaborate theory. It will turn out that the CQT faces serious problems in the light of quantum physics. Nevertheless, none of the problems, in my opinion, are knock down arguments and finding a reply will lead the way for the following chapters.

2.2 A short history of transference theories

Maybe the first to propose that causation is a physical process and has something to do with conserved quantities was Aronson (1971).[1] His motivation to develop such a theory comes from the problem of how we can account for the apparent asymmetry of the causal relation, especially in cases of simultaneous causation. For Aronson, pointing to the asymmetry of manipulation (we can only manipulate the future, but nor the past) is no solution, since this would impose an anthropomorphic concept on physical processes, which have nothing to do with human agency (see my discussion in section 4.3). For him "the challenge is to find a basis for determining a

[1]Even though the idea to reduce causation so something physical is undoubtedly much older (cf. e.g. Ducasse 1926).

cause-effect asymmetry without having to appeal to human manipulation." Aronson (1971, p. 416) After having argued that only dynamical changes are instances of causation, Aronson wants to achieve this goal by taking the following as a condition for causation:

> Prior to the time of the occurrence of B [the effect], the body that makes the contact with the effect object possesses a *quantity* (e.g. velocity, momentum, kinetic energy, heat, etc.) which is transferred to the effect object (when contact is made) and manifested as B. (Aronson 1971, p. 422)

Even though in this quote Aronson is indefinite as to which quantity is transferred in causal interactions, his later emphasis on conservations principles, which secure that quantities do not come into existence out of nothing, suggests that Aronson actually only means conserved quantities, which would eliminate velocity and heat from his list.

An immediate criticism of Aronson's account is that he fails to deliver what he sets out for, namely to provide an explanation for the asymmetry of causation. As will become clear in chapter 4, simply pointing to physics does not solve the problem and even though it might be the case that physicists occasionally use asymmetric notions like 'transfer', any theory of causation has to be established on the ground of the mathematical formalism or models thereof, where asymmetry plays no obvious role. As a result, I regard Aronson's proposal merely as a first and preliminary formulation of the core idea, and will now go on to the next step that has been made in the development of physical causation.

Fair (1979) picks up the idea of Aronson, though he has a slightly different motivation in that he suggests "taking seriously what physical science has discovered in answering the traditional philosophical questions about the nature of causation." (Fair 1979, p. 228) Leaving aside whether physics is able to discover the nature of causation, Fair is right in criticising Aronson for being unclear which properties are transferred by the causal relation, in particular he holds the transfer of force to be inconsistent with physics. Instead, Fair regards causation only as the transfer of energy or momentum, even though he does not give any reasons for neglecting other physical properties like charge. Furthermore, Fair believes that energy and momentum are not always transferred together, but that there are processes in which only the transfer of one of them constitutes the causal relation. His example for such a process is a charged particle in an electromagnetic field for which, as he maintains, only the direction of its momentum changes, but not its energy.[2]

Another point, which Fair criticises, is Aronson's view that the exact amount of a quantity that is lost by the cause in a causal interaction has to be gained by the effect. According to Fair this would exclude processes in which some energy is lost due to dissipative effects from being causal. Therefore, only some amount of the energy or momentum, lost by the cause, has to be transferred to the effect in order

[2] This, of course, is not quite right, since a charged particle in an electromagnetic field will also change its potential and kinetic energy.

to establish a causal relation.[3] On the other hand, Fair maintains that it is the exact same energy or momentum that is first lost by the cause and then gained by the effect. For Fair, the identity conditions of properties, which are necessary to establish this claim, are constituted by conservation laws. He argues that if energy and momentum are conserved in an area, comprising cause and effect, then we can be certain that energy and momentum retain their identity over time, because the conservation laws forbid that conserved quantities leave or enter this area. I believe that this argument is not without problems and will discuss it later in more detail.

Despite these differences, Fair and Aronson agree that physics can ground the asymmetry of causation, simply by pointing out that we can define the direction of the transfer of energy-momentum by the direction of the flow of conserved quantities through a surface surrounding the cause. The problem that this definition seems arbitrary in the light of time reversal invariant physical laws is not discussed by Fair. Furthermore, Fair only briefly mentions a concern raised by Earman (1976), namely that the causal relation so defined is not the same for all reference frames. Fair gives the example of one billiard ball hitting another, which looks very different from the reference frame of the table compared to the reference frame of one of the balls. However, Fair sees no difficulty here and acknowledges the relativity of the causal direction to reference frames.[4]

Following the evolution of physical causation chronologically upwards, the next in line is Heathcote (1989). According to Heathcote the previous attempts to define physical causation are insufficient, because first they are not specific enough about what quantities are transferred, second they leave unclear whether they seek to establish causation only for the macroscopic or the microscopic level or both, and third they are not relativistically invariant. On the contrary, Heathcote starts from the observation that if we want to establish a theory of physical causation, it is QFT where we have to look for it, because only QFT, not Newtonian physics or QM, describes the interaction between particles. Heathcote (1989, p. 83) identifies three conditions that have to be fulfilled by physics to make it compatible with causation and avoid causal paradoxes: (1) 'Stable Causality of Space Time' which means that the topology of spacetime should not form loops, (2) 'Locality principle of causality in accord with the Special Theory of Relativity' that is, nothing travels faster than the speed of light and (3) 'Time Ordering principle of causality' which says that effects always follow their causes in time. In particular the second condition, Heathcote argues, is realised by the commutator relations of QFT and the third condition is at least part of the Wightman axioms of axiomatic QFT (AQFT).

On the other hand, Heathcote (1989, p. 91) believes that as a consequence of Haag's theorem, one has to regard interactions in QFT as a 'black box'. This belief keeps him from exploring further the details of interactions in the Lagrangian formulation of QFT. Instead, he turns towards AQFT, a theory that is unable

[3]This critique on Aronson, however, seems to be unfair. Aronson could reply that also in his account dissipation can be described as energy transfer and interactions involving dissipation as a causal interaction with several different effects.

[4]I will discuss these problems in detail in chapter 4.

to describe interacting fields. This stance was perfectly reasonable at the time Heathcote published his paper, however, since then the discussion has evolved. As I will explain in section 6.2, today interactions can be described in the Lagrangian formalism despite Haag's theorem. I find it slightly puzzling, how Heathcote intends to do justice to his main idea that QFT as a theory of interactions is particularly suitable for causation and at the same time relies on AQFT, but be that as it may. He helps himself by retreating to a more phenomenological description of interactions that is merely inspired by the mathematical formulation, and which I will briefly present in the next paragraph.

Heathcote adopts an ontology that consists first of leptons and quarks that he understands as particles, i.e. "pointlike masses" (Heathcote 1989, p. 100), and second of bosons that are field quanta, transmitting interactions and sharing only some characteristics with particles. The latter means that he is a realist about so-called virtual particles. With this at hand, Heathcote puts his version of the conserved quantity theory into the statement: "[A]ll causal influences are the result of forces between objects, all such forces are interactions in the sense of QFT." (Heathcote 1989, p. 101 f.) Hence, for example, charged particles, like electrons, interact by exchanging virtual particles, namely photons. Like Fair, Heathcote sees energy and momentum (and angular momentum) as the only properties that are exchanged in causal interactions. His reason for holding this view is that all other properties "are fixed and do not change as a result of causal influence." (Heathcote 1989, p. 104) Furthermore, Heathcote stresses the connection between transference theories and conservation laws by explaining that whenever one particle changes its state, conservation laws necessitate that this change has to be compensated by a change in another particle and, according to QFT, this cannot happen by action at a distance, but only via the exchange of virtual particles. He even goes so far as to "say that causality is exchange of quanta of a field because the conservation laws exist." (Heathcote 1989, p. 104)

2.3 Salmon's process theory

Salmon's process theory of causation drops a bit out of line from the other theories I am discussing in this chapter. This is for two reasons: First, he does not rely only on conserved quantities being transferred and second, he puts the emphasis on causal processes, rather than interactions. Both characteristics, however, do not make Salmon's theory incompatible with the others. As will become clear, Salmon's criterion of 'mark transmission' can be understood as transmission of a conserved quantity and conversely transference theories can be constructed as process theories of causation.

By turning towards processes as basic constituents of causation, Salmon wants to dissociate himself from the tradition that conceives causation as a relation between distinct events: "It is my conviction that this standard view, in all of its well known variations, is profoundly mistaken, and that a radically different notation should be developed." (Salmon 1984, p. 138) The boundaries, however, between events and processes are left unsharp by Salmon; events are fairly well localised in spacetime, while processes typically extend over much larger temporal than spatial regions (cf. Salmon 1984, p. 138). More important for him seems to be how we can distinguish causal from pseudo processes. The basic criterion to do this is the following: "A causal process is capable of transmitting a mark; a pseudo process is not." (Salmon 1984, p. 142) A classical example for this distinction is an object, like a ball, that one can mark, e.g. by scratching it, but one cannot mark the shadow of the ball. More precisely, a causal process has to have a certain structure and retain this structure over time, because only then can we mark it by modifying the structure (cf. Salmon 1984, p. 144). Furthermore, causal processes have to behave according to the laws of physics, pseudo processes not (cf. Salmon 1984, p. 142 ff.; 179). However, Salmon explicitly dismisses using conserved quantities as the criterion, because they cannot distinguish between "situations in which energy is transmitted from those in which it merely appears in a regular fashion." (Salmon 1984, p. 146)[5] Putting all this in a single statement, Salmon is able to define the criterion of 'mark transmission':

> MT: *Let P be a process that, in the absence of interactions with other processes, would remain uniform with respect to a characteristic Q, which it would manifest consistently over an interval that includes both of the space-time points A and B (A $\neq$ B). Then, a mark (consisting of a modification of Q into Q'), which has been introduced into process P by means of a single local interaction at point A, is transmitted to point B if P manifests the modification Q' at B and at all stages of the process between A and B without additional interventions.* (Salmon 1984, p. 148)[6]

Using this definition of causal processes, Salmon goes on to define the propagation of causal influence:

> PCI: *A process that transmits its own structure is capable of propagating a causal influence from one space-time locale to another.* (Salmon 1984, p. 155)

In other words, a causal process can be marked, propagate this mark and mark other processes in causal interactions. Referring to Hume's critique of causation, Salmon expresses the conviction that the "propagation of causal influence by means of causal processes constitutes, I believe, the mysterious connection between cause and effect which Hume sought." (Salmon 1984, p. 155) To understand this claim, it has to be read in the context of Salmon's understanding of processes in the frame of an "at-at"

[5]In my opinion, Salmon's worry is unjustified. A moving light spot on a wall is not governed by a conservation laws and can simply vanish without violating energy conservation. A causal process, on the other hand, that carries energy cannot vanish without the energy being transferred somewhere else. Since Salmon holds that causal processes are governed by the laws of nature, embracing conserved quantities would be a small step for him.

[6]It might be noted that MT, as well as the later definition CI, contains counterfactual statements and is therefore problematic, since it is not always clear how these should be evaluated. Salmon essentially holds that the truth of counterfactuals can be found out by empirical experiments (cf. Salmon 1984, p. 148 ff., but a critique can be found in Dowe 2000, ch. IV.3.4.).

theory of propagation (cf. Salmon 1984, p. 152 f.). That means, every process is completely described by being *at* a certain point in space *at* a certain time and it is not a sensible question to ask what happens in between these points. With respect to Hume's question, what happens between two billiard balls when they meet, there is then nothing more to say that both balls meet at a certain point in space at a certain time; there is no question of what happens in between the balls, that is, in between different 'at-at' points.

The last thing that needs to be defined, in order to complete Salmon's account, are causal interactions, that is, how a causal influence can be transmitted from one process to another. The definition is the following:

> CI: *Let P_1 and P_2 be two processes that intersect with one another at the spacetime point S, which belongs to the histories of both. Let Q be a characteristic that process P_1 would exhibit throughout an interval (which includes subintervals on both sides of S in the history P_1) if the intersection with P_2 did not occur; let R be a characteristic that process P_2 would exhibit throughout an interval (which includes subintervals on both sides of S in the history of P_2) if the intersection with P_1 did not occur. Then, the intersection of P_1 and P_2 at S constitutes a causal interaction if:*
>
> (1) *P_1 exhibits the characteristic Q before S, but it exhibits a modified characteristic Q' throughout an interval immediately following S; and*
>
> (2) *P_2 exhibits the characteristic R before S, but it exhibits a modified characteristic R' throughout an interval immediately following S.* (Salmon 1984, p. 171)

Accordingly, it is essential for causal interactions that two or more processes meet in spacetime and that at least one property of each one undergoes a change that remains with the process after the interaction. Furthermore, CI is a necessary and sufficient condition for causal interactions, if it is assumed that there are properties such as Q and R and that they fulfil CI. In short, Salmon's process theory can then be formulated by using the concepts of propagation and production: "Causal processes are the means by which causal influence is *propagated*, and changes in processes are *produced* by causal interactions." (Salmon 1984, p. 170) However, this process theory was effectively criticised by Dowe (1992, 1995), which led Salmon to change his theory to something that is very close to what I am going to discuss next.

2.4 Dowe's conserved quantity theory

Dowe developed his conserved quantity theory out of a discussion of transference and process theories, which have been introduced in the last two sections, and he took elements from both of them. Roughly, from Aronson and Fair he adopted conserved quantities, and from Salmon he adopted processes. I consider Dowe's conserved quantity theory to be the most elaborate theory of physical causation today and will now present it in detail.

Dowe characterises the relation of his theory to that of Salmon in the following way: "The approach to be taken is to modify Salmon's theory by introducing the concept of a *conserved quantity*." (Dowe 2000, p. 89) He then sees his task in answering three main questions:

1. "what are causal processes and interactions?"
2. "what is the connection between cause and effect?"
3. "what distinguishes a cause from its effect?" (Dowe 2000, p. 89)

The answer to the first question is given by Dowe in two definitions:

CQ1. A *causal process* is a world line of an object that possesses a conserved quantity.

CQ2. A *causal interaction* is an intersection of world lines that involves exchange of a conserved quantity. (Dowe 2000, p. 90)

In order to understand the definitions a few explanations are needed. With respect to CQ1, first of all a *world line* is the trajectory of an object in a Minkowski diagram. Dependent on whether the object is point like or extended in space, the world line will either be only a line, that is, at every time instant only one point in space is occupied by the world line, or the world line will be an 'elongated region' and on a timelike hypersurface of spacetime the world line will occupy more than one point. World lines will therefore differ corresponding to whether the ontology contains, e.g., only point particles or fields. Dowe remains very liberal when it comes to the question of what kinds of things exist, and takes an *object* to be whatever there is according to science. Apparently his opinion is that different ontologies will not make a difference to his account of causation. Though, a restriction on the range of objects is that they have to retain their identity over time. A further restriction Dowe introduces is that objects must possess a *conserved quantity*, that is, every property that is subject to a conservation law in physics. As a result, CQ1 should be sufficient, as Dowe maintains, to sort out causal processes from non-causal pseudo processes (cf. Dowe 2000, p. 90 f.).

With respect to CQ2, Dowe defines an *intersection* as the meeting of world lines in a Minkowski diagram. Such a meeting can consist of two or more world lines and occupies the spacetime region in which at least two of the world lines overlap. Finally, an "*exchange* occurs when at least one incoming, and at least one outgoing process undergoes a change in the value of the conserved quantity, where 'outgoing' and 'incoming' are delineated on the spacetime diagram by the forward and backward light cones, but are essentially interchangeable." (Dowe 2000, p. 92)

A causal process as defined by CQ1 exemplifies what Dowe calls 'causation as persistence'. His example is a spaceship moving uniformly in space and where, according to Newtonian mechanics, the only cause for the motion is the ship's own inertia. Dowe's motivation to assume that this case has anything to do with causation is its "inherent plausibility" (Dowe 2000, p. 53). An example for a causal interaction, as defined by CQ2, on the other hand is the Compton effect, that is, the interaction between light and matter. In a (classical) description, the cause is an electron that possesses a certain energy and then is hit by a photon, viz., both world lines overlap. The resulting effect is the photon at a later time possessing a smaller energy than it did before the interaction.

Now the answer to the second question, regarding the connection between cause and effect, has to be given. This is important, since intuitively there can be a causal connection between two objects even if their world lines do not meet, but are merely connected by a third world line. However, Dowe first notes that the connection cannot simply be established by any causal process that is linked to a cause and an effect. The main problem is that this would allow for misconnections. For instance, in a physical laboratory a ceiling light shines down on an apparatus that measures the outcome of an experiment. There is a causal process, namely photons, that link the light and the apparatus, but the light is typically not the cause for what the apparatus measures (cf. Dowe 2000, ch. VII.1). Being connected by a causal process is certainly a necessary condition for causal connection, but as the above example shows, it is not sufficient. Hence, Dowe tries to combine connection via causal processes with the condition that the cause raises the chance of the effect. However, after an extended discussion he dismisses this combined theory for two reasons. First, it cannot sufficiently handle examples of chance lowering causes and second, what one might call a version of the problem of preemption, there might be cases in which two objects are linked by a causal process and one object raises the chance for the occurrence of the other, but still does not cause the other.

The solution that Dowe ultimately sees for the problem of misconnections is to refine the notion of being linked by a causal process. He begins by taking facts to be the relata of a causal relation. This however is not much more than a simplification of terminology, because a fact "is an object having a property at a time or over a time period." (Dowe 2000, p. 170) Dowe then states his definition of a causal connection:

> *Causal Connection*: There is a causal connection (or thread) between a fact q(a) and a fact q′(b) if and only if there is a set of causal processes and interactions between q(a) and q′(b) such that:
>
> 1. any change of object from a to b and any change of conserved quantity from q to q′ occur at a causal interaction involving the following changes: Δq(a), Δq(b), Δq′(a), and Δq′(a) (sic!); and
> 2. for any exchange in (1) involving more than one conserved quantity, the changes in quantities are governed by a single law of nature. (Dowe 2000, p. 171f.)

These conditions solve the problem of misconnections by exactly specifying what the conserved quantities are and what amount of them is exchanged. The example given above, of the ceiling light shining on the measurement apparatus, is then ruled out, because the energy that the photons of the light exchange with the apparatus is in general not of the right amount to cause the measurement outcome. Furthermore, condition (2) is needed to "rule out cases where independent interactions occur by accident at the same time and place." (Dowe 2000, p. 172)

An example of a causal relation, where cause and effect have no overlapping world lines, but are connected by a causal process, and where additionally not only one conserved quantity is exchanged, is the following. A photon hits an atom and is absorbed. As a result, the atom decays to a different atom with different charge. The cause is photon a with energy q at time t_1 (in the notation of Dowe: $q(a)$ at t_1). The first causal interaction is the intersection of photon a with atom b, in which a exchanges energy with b $(\Delta q(a), \Delta q(b))$. The second interaction is the decay of atom b to atom c, in which energy q and charge q' are exchanged $(\Delta q(b), \Delta q(c), \Delta q'(b), \Delta q'(c))$. Following the definition of a causal connection, the effect is atom c with charge q' at time t_2 ($q'(c)$ at t_2) (cf. Dowe 2000, p. 172 f.).

I will not discuss Dowe's answer to the third question *'what distinguishes a cause from its effect?'* here, given that I present my own very different account for that in chapter 4. Dowe's theory, which can be found in chapters seven and eight of Dowe (2000), is based on the supposed existence of backwards causation and on Reichenbach's asymmetry of causal forks (cf. Reichenbach 1956). Price (1996a) has presented a forceful criticism of Dowe, with which I ultimately agree. However, Dowe's definitions of causal interactions and causal connections can be retained without his account of causal asymmetry, and for the present work I take them to be the take-away message of the CQT.

2.5 Discussion

Having introduced various theories of physical causation that all share the basic thought that causation is some property being transferred in causal interactions, I now wish to discuss some points that I regard to be problematic.

To begin with, the accounts of Aronson, Fair and Salmon have been sufficiently criticised by Dowe, so I can only repeat his points here (cf. Dowe 2000, ch III.4). Dowe sees two main failures of Aronson and Fair, the first is that both argue for the possibility to ascribe an identity to conserved quantities. However, the simple appeal to conservation laws is not enough to underpin this claim, since the latter only guarantee numerical equivalence. It is therefore not justified to talk of a transfer of conserved quantities, but merely the exchange can be retained, where of course the term 'exchange' is more or less stipulated to have a meaning that does not involve identity of properties. Dowe's second critique is that Aronson and Fair cannot explain the asymmetry of the causal relation, because, as will become clear in section 4.2.3, the appeal to temporal asymmetry alone is not sufficient.

Dowe also disagrees with Aronson and Fair over which conserved quantities are exchanged in causal interactions. It is obvious that the vagueness of Aronson, who does not exactly specify which properties he means and also includes unconserved quantities, is unacceptable, but Dowe is also more liberal than Fair. While the latter sees only energy-momentum as being transferred, Dowe cannot find any problem in letting every conserved quantity play that role. However, neither does Fair give an argument for his restricted view, nor does Dowe for his liberal view. Here one can invoke Heathcote who, as already mentioned above, takes the side of Fair and argues that only energy-momentum is exchanged, because other properties do not change their values as a result of interactions. Here I agree with Heathcote, but this is a point that will be evaluated further in chapter 7.

With respect to Salmon's process theory, Dowe finds its most severe failure in its circularity. In short, the problem is that the concepts of 'mark', 'mark transmission' and 'interaction' are mutually dependent. On the one hand, a mark can only be made by a causal interaction and only be transmitted by a causal process in the absence of causal interactions. On the other hand, an interaction is defined such that a mark is transmitted after the interaction has occurred. Consequently, as a complete definition of causation, Salmon's process theory is circular and thus not tenable.

This critique led Salmon to revise his theory and take over large parts of Dowe's (1992) version of the CQT, with the difference, however, that Salmon relies on invariant quantities, rather than conserved quantities. Invariant quantities are those that do not change their value in different inertial frames, in contrast to conserved quantities that may vary from one inertial frame to another. Salmon's motivation for this restriction is that causal interactions should be invariant between inertial frames (cf. Salmon 1994, ch. 7). However, Dowe (2000, p. 119) convincingly argues that, even though the values of conserved quantities might change in different inertial frames, the change of a conserved quantity is in fact invariant. Therefore, Dowe's notion of exchange of conserved quantities is also invariant and no disagreement between different inertial frames can arise as to whether a causal interaction takes place or not.[7]

Coming now to Dowe's own theory, he claims to have solved the problem of the identity of properties that is contained in the notion of 'transfer' by using the weaker 'exchange'. However, Kistler (2006) argues that this is in fact not a possible route to take. Instead, Kistler wants to keep the notion of 'transfer' for the reasons that "we can justify the thesis that the causal relation is objectively more than mere 'contiguous succession' only if it is possible to justify the idea that the same amount [of a conserved quantity] can appear in two distinct events." (Kistler 2006, p. 53) So the problem for Kistler is that there is nothing that distinguishes the CQT from a usual regularity theory of causation if conserved quantities are not transferred. In other words, if it is not the case that the identical conserved quantity, that is lost by the cause is gained by the effect, then the CQT does not describe causally connected events, but merely the succession of events. However, I cannot agree with Kistler here, because first of all he does not show how the notion of 'transfer' could be justified, but more importantly, giving up transfer does not transform the CQT into a usual regularity theory. The reason is that the CQT gives criteria for the identification of causal connections that are stronger than what is given by regularity theories. Even though Dowe's criterion that world lines have to meet is analogous to 'continuity' criteria of regularity theories, the further criterion that the amounts

[7] To be sure, there is more disagreement between Dowe and Salmon than I present here (cf. Salmon 1997; Dowe 2000, ch. V.6), but following their discussion in all its details is beyond the scope of this chapter and, in my opinion, without much merit for my overall project.

of conserved quantities, possessed by the meeting objects, have to change is an additional criterion that sorts out causally related events from events that only stand in 'contiguous succession' to one another.

The next problem I want to discuss is the threat of circularity. It has been criticised that the CQT is circular, because what a conserved quantity is can only be defined by invoking causation: "A conserved quantity is one that remains constant through time in a closed system, but what is a closed system but a system that does not engage in any causal interaction?" (Hitchcock 1995, p. 315 f.) Dowe responds by arguing that for instance energy can be defined in a different way: "energy is conserved [...] on the assumption that there is no net flow into or out of the system." (Dowe 2000, p. 95) However, it is unclear whether 'net flow' is something other than a causal process according to the definition CQ1 (cf. Schaffer 2001, p. 810). Nevertheless, in my opinion, this critique is ineffective, since the CQT is not an analytic definition of the concept of causation. If it were, of course, conceptual circularity would be severe. Instead, the CQT points at processes in the world that can be understood as causal and this can be done without caring about circularity. If we want to know which quantities are conserved, we do not have to use an analytic definition of what the concept 'conserved quantity' means, but we have to look into a physical theory, which has already done the work of finding these quantities.

Another problem for the CQT stems from the notion of causal processes. Dowe (2000, p. 90) defines a causal process as "a world line of an object that possesses a conserved quantity" and justifies the description of processes as causal via causation as persistence, that is, an object being the cause for its own state in the future. However, understanding processes as a series of causes and effects seems superfluous. The only example that Dowe gives for a causal process is a case where the inertial mass of an object is the cause for its uniform motion. Against the interpretation of this as causal speaks the theory of special relativity according to which uniform motion in one inertial frame is rest in another one. Consequently, I do not see any reason why uniform motion should be in need of a causal explanation, if we can simply erase the causal process by changing the reference frame. The same goes for persistence of an object in general. If we take the very basic intuition that causes make a difference to their effects, then it is unclear in what sense an object makes a difference to itself by persisting in time. Having said that, giving up causal processes does not seem to threaten the other parts of the CQT. We still have the definition of causal interaction and causal connection that remain applicable even if causal processes are not causal anymore. Of course, one might still call a process causal, if it carries a conserved quantity and is involved in causal interactions; all I am arguing against is that processes are a series of causal interactions between earlier and later stages of itself (see also my discussion in section 3.3.4).

I will now turn to issues that are related to quantum physics. Here the main concern is that the notion of world lines becomes highly problematic. The Heisenberg uncertainty principle tells us that an object cannot have a sharp position and a sharp momentum at the same time. Additionally, states in quantum physics do not have to be pure states but can be superpositions. So there are objects that do not have a sharp value of any property. This seems to make it impossible to define a sharp world line of an object (more on this in chapter 5), and in turn makes Dowe's notion of "exchange" opaque, for exchange can hardly be defined as the intersection of world lines, if world lines have no sharp intersections.

Second, as already mentioned, following Price (1996a), Dowe's explication of the directionality of temporal causation, which to a large extend relies on backwards-in-time causation as an explication of measurement outcomes on entangled states, is highly problematic. Due to restrictions of space I cannot present Dowe's proposal here. However, in chapter 4 I will analyse the problem of the directionality of causation in detail and offer a way for solving it.

A more general worry about the CQT might be that, given Dowe's motivation to deliver a theory of causation based on physics, it has a surprisingly poor connection to the latter. In particular, Dowe does not mention QFT even once and hence does not consider the question whether the CQT and QFT are compatible. As we will see, the already mentioned problem of the definition of world lines and their intersections becomes even more inescapable in QFT. Additionally, Dowe defines causation as a relation between objects, where objects are what exists according to the theory. However, there is currently no sufficiently established ontology for QFT, and therefore the content of the CQT is underdetermined. These two concerns will be some of the main problems discussed in chapter 5.

Heathcote (1989) is not much help here, even though his goal is explicitly to show what causation is in QFT. He leaves it with the vague statement that causation is interaction as defined by QFT and that energy-momentum is exchanged by virtual particles between interacting particles. This is highly problematic for several reasons. First, the status of virtual particles is unclear in a way such that their existence cannot be taken for granted (more on this in section 5.3.5). Furthermore, the problem of virtual particles is connected to the question what we should make of perturbation theory, whether it is only a mathematical tool or more than this. The status of perturbation theory in turn is linked to the problem whether we can see QFT as a temporal description of a dynamical process or whether it just describes asymptotic initial, final states and transition probabilities, but not what happens during the interaction (more on this in section 6.2). Additionally, Heathcote interprets leptons as particles, in a quite literal sense, and bosons as field quanta, but again these ontological interpretations of QFT are dubious. Nevertheless, one might learn from Heathcote, that Dowe's definition of causal interactions as intersections of world lines is a too simple picture. Heathcote highlights the importance of certain general principles in QFT, such as locality, and maintains that they are helpful for a causal interpretation. Even though I believe Heathcote's analysis is incomplete, I do agree with his approach and spell out the physical basis in more detail in chapter 5.

2.6 Conclusion

All accounts presented in the above sections agree on the basic assumption that causation is the transfer or exchange of some physical property from cause to effect. This might be taken as an indication that this theory of causation is on the right track. However, when diving into the murky details this becomes much less clear. The theories of Aronson, Fair an Heathcote are not sufficiently developed to see where the problems are. On the contrary, Dowe took the necessary step and spelled out his theory in much more detail, but by doing so, opened up many possibilities for disagreement with physics. In particular, Dowe's ignorance of QFT is unacceptable given its importance, and invoking it puts the applicability of the CQT in question. Dowe's use of world lines and their intersection becomes questionable in the context of QFT and, second his use of objects as causal relata, where this is cashed out according to the ontology of a certain theory, makes it look like we have to wait for a suitable ontology for QFT to come up before the CQT can be further developed. It should also has to be mentioned that Dowe's account for the asymmetry of causation is hardly convincing and asymmetry is still a problem that needs to be solved.

In short, I take the agreement of various authors about what causation is physically as a motivation to investigate whether this idea can be rescued in the light of quantum physics and the problems that have surfaced as a guide to what needs to be done.

3 The meaning of ‘causation’

3.1 Introduction

In the last section I have discussed the various proposals for theories of physical causation that have already been put forward. The result was that none of them could sufficiently account for the challenge that QFT poses, that is, none of them can be applied to QFT, and identify processes in QFT as causal or not causal. This means that these theories of physical causation cannot live up to their own expectations; their promise and goal was to deliver a theory of what causation is in the physical world, but they fail to pay sufficient attention to one of the most fundamental physical theories today. I want to suggest that this failure cannot simply be resolved by making adjustments to the existing theories of physical causation. The main problem with this strategy is that prima facie it is unclear where changes can be made to which extent, and in particular it leaves open when changes actually go too far and turn a theory of physical causation into a theory that is not about causation anymore. There is a point at which changes to a physical theory of causation are so severe that one is not talking about causation anymore after the changes are done, and we need to know where that point is. What is required to not fall into this trap is a systematic approach to causation.

Another motivation comes from the fact that different philosophers have understood very different things under the term ‘causation’. For example, David Hume found that what we, i.e., the folk, usually mean by the term ‘causation’ implies that there is a necessary connection between cause and effect. He then continued to reject such an account of causation on the grounds that the concept of ‘necessary connection’ does not make any sense. This opens up several questions for consideration: First, assuming that one has the task to find a theory of causation that is coherent with what the folk think about it. Does this mean that one has to incorporate ‘necessary connection’ into one’s theory, even though it is supposedly senseless? Second, there are certainly philosophers who believe that they can make sense out of necessary connections, e.g., some dispositionalists. As a consequence, can they use necessary connections to analyse ‘causation’? And third, there are other philosophers, who believe that causation is not a necessary connection for completely different reasons than Hume, for example because they believe that causation is the relation of chance raising. Which argument can they make to counter someone who thinks that causation is a necessary connection, because that is what we commonly think it is? Again, these problems are in the first place not about what causation is, but related to the method of investigation, and again they motivate a systematic approach to the question of what causation in physics is.

In conclusion, what we need before going into physics are criteria that tell us what counts as a theory of causation and what not. Only on this basis can processes in QFT be identified as causal in a non ad hoc way. However, if we start the investigation from scratch, this is not quite enough. Before establishing such criteria it needs to be clear how they can be found, and on which grounds criteria have to be accounted for or can be dismissed in the final theory. Thus, the first step for the analysis of causation has to be to establish a proper method for doing so. Accordingly, I will start this section by going into methodology. I will discuss the most prominent methodological paradigms today: on the one hand conceptual analysis as following the Canberra plan, and the naturalistic spirited empirical analysis on the other hand. The conclusion will be that, at least for present purposes, both methods coincide to an extent where they are practically identical. For both the first step is to find the platitudes of causation, or in other words, a folk theory of causation. The next step then is to find out whether anything in the world realises these platitudes. Therefore, following this method, I will proceed in the subsequent sections by trying to establish an analysis of causation in terms of a collection of platitudes. The second step, that is, going into physics, will be the topic of the subsequent chapters. If successful, the benefits of this method are twofold. First, by analysing the opaque concept of causation into better understood platitudes, one gains a superior understanding of the concept. Second, if it turns out that the platitudes are realised by physics and thus that there is causation in physics, our understanding of physics is improved.

3.2 Methodology: The Canberra plan vs. naturalism

In this section I want to briefly explain the methodological basis for this investigation into causation. The two main paradigms in methodology today are conceptual analysis, usually in the style of the Canberra plan, and naturalism. I will start this section by introducing the Canberra plan and by explaining how it can be employed for the concept of causation. I will then go on to the rival methodological account of naturalism, and after an exposition, again show how naturalism handles the case of causation. In the final part of this section I will draw a comparison between the Canberra plan and naturalism. The most important conclusion will be that, at least in the case of causation, there is now significant difference between the two methodological accounts. Both start off by finding platitudes about causation and then go on to look for what realises these platitudes in the world. The only noticeable distinction is the status of the platitudes. While Canberra planners usually hold the platitudes to be a priori truths, naturalists regard them as contingent beliefs, historically formed and possibly dependent on cultural background or other social influences. However, as I will argue, for two reasons this disagreement is sufficiently weak to be without impact for my research on causation. First, the concept of the a priori among Canberra planners is far enough removed from its original meaning to be nearly compatible with naturalism, and second, since a decision in this dispute is of no consequence for the methodological procedure and since the aim of my work is on causation, not the status of the a priori, I can happily live without ultimately taking sides on this matter.

The Canberra plan has its origin in Lewis (1970, 1972), who in turn relies on the works of Ramsey and Carnap. Lewis is concerned with two scenarios, namely, that of an old theory which is replaced by a new one, and that of a theory which is reduced to another. In both cases Lewis presents a strategy for how the vocabulary of one theory can be expressed through the vocabulary of the other. Frank Jackson developed Lewis' proposals further and propelled them to a general method for inquiries into whether a certain thing, say a mental state, exists. If one wants to answer such questions, one first of all has to know the meaning of the concept, the existence of which is at stake. For Jackson, to spell out the meaning of a concept is to find our intuitions about it. Ideally, the intuitions should form a coherent body, which can be called the theory of a concept, and should coincide with the intuitions of most people. In short, "your intuitions reveal your theory. To the extent that our intuitions coincide, they reveal our shared theory. To the extent that our intuitions coincide with those of the folk, they reveal the folk theory." (Jackson 1998, p. 32)

The process of finding what a concept means by exposing our intuitions about it, is what Jackson calls conceptual analysis. The latter can be given a modest or an immodest role in philosophy. In its immodest role, conceptual analysis not only gives the meaning of a concept, but actually tells us what the world is like. On this view, the fact that we have a certain concept with a certain meaning implies that there is something in the world that realises this concept.[1] Nevertheless, this is not how Jackson understands conceptual analysis:

> [T]he role for conceptual analysis that I am defending in these lectures is the modest role: the role is that of addressing the question of what to say about matters described in one set of terms given a story about matters in another set of terms. Conceptual analysis is not being given a role in determining the fundamental nature of our world; it is, rather, being given a central role in determining what to say in less fundamental terms given an account of the world stated in more fundamental terms. (Jackson 1998, p. 44)[2]

On the other hand, for Jackson conceptual analysis is to spell out a priori knowledge. This claim rests in a specific theory of meaning, which I cannot go into here. What is important is that the notion of the a priori that Jackson defends is very weak, in that it does not imply that a priori knowledge is completely detached from everything empirical. It is first of all compatible with a theory of representation that regards

[1] This is how Norton (2007, p. 15) sees what he calls causal fundamentalism, that is, "as a kind of *a priori* science that tries to legislate in advance how the world must be." I agree with Norton that this is not an appropriate role for metaphysics, and this is not the role that metaphysics has in the present work.

[2] Ladyman & Ross (2007, p. 16) criticise Jackson by asking "why should we think that the products of this sort of activity [i.e. conceptual analysis] reveal anything about the deep structure of reality, rather than merely telling us about how some philosophers, or perhaps some larger reference class of people, think about and categorize reality?" Given that Jackson endorses only the modest role of conceptual analysis, the critique by Ladyman and Ross seems to rest on a misunderstanding of Jackson's methodology.

intensions and extensions to be causally or historically generated. Furthermore, results of the conceptual analysis are fallible and subject to possible changes over time. Jackson’s conception of the a priori is thus far away from more traditional accounts, which are closer to Kant (cf. Jackson 1998, p. 56).

When the conceptual analysis is finished, one can proceed to the next step in the Canberra plan, that is, to investigate whether there is something in the world that realises our intuitions about a certain concept. This question has to be answered empirically, and thus commonly depends on the natural sciences. It can therefore not come as a surprise that most Canberra planners are physicalists and only accept physical realisers (cf. Braddon-Mitchell & Nola 2009, p. 8). In short Nolan (2009, p. 268) describes the complete method of the Canberra plan as follows: “We select something we would like a philosophical analysis of: causation, color, free will, beliefs, moral value, or whatever. Then we collect together the ‘platitudes’ concerning our subject matter. [...] Once we have these platitudes assembled, we conjoin them, and use the resulting information to define a ‘theoretical role’ for the thing we are interested in. [...] At the second stage, we look at our theory of the world to tell us what, if anything, plays the role so defined.” Even though much more could be said on the Canberra plan, I will now turn towards naturalism.[3]

The term ‘naturalism’ has different meanings in different areas of philosophy, however in the area of methodology there are some core ideas shared by most if not all naturalists, which can be regarded as the orthodox method in Anglo-American philosophy today (cf. Caro & Macarthur 2004, p. 1). This core can be put into two propositions, though again, there exist many variations of them:

> 1. *An Ontological Theme*: a commitment to an exclusively scientific conception of nature; 2. *A Methodological Theme*: a reconception of the traditional relation between philosophy and science according to which philosophical inquiry is conceived as continuous with science. (Caro & Macarthur 2004, p. 3)

According to the first theme, only science provides a true picture of the world. Usually, ‘science’ here means physics, and thus the first theme amounts to a physicalistic ontology. The second theme, on the other hand, amounts to a denial of the possibility of a First Philosophy, which would rest on a priori knowledge. Consequently, philosophy has no authority to judge the claims of science and has no means to find knowledge independently of science (cf. Kitcher 1992, p. 74-76; Caro & Macarthur 2004, p. 6). This side of naturalism is often divided into a descriptive and a normative part (cf. Quine 1969; Devitt 1984, p. 76). In the descriptive part it is studied how we acquire knowledge within science. In the normative part the justification for this knowledge is given. My emphasis here is on the justificatory part, which can be expressed as follows:

> The *methodological (or epistemological) scientific naturalist* holds that it is *only* by following the methods of the natural sciences – or, at a minimum,

[3]It might be noted that ironically Lewis (2004a) has argued against the applicability of the Canberra plan to causation. However, see Liebesman (2011) for a convincing reply.

> the empirical methods of a posteriori inquiry – that one arrives at genuine knowledge. (Caro & Macarthur 2004, p. 7)

In other words, all knowledge is in the end justified by science, in that only science shows how justified knowledge can be found empirically. Even knowledge that many people intuitively would regard as a priori, e.g., mathematical or logical knowledge, can only be justified empirically, an assumption that results from the Duhem-Quine-thesis of justificatory holism (cf. Devitt 2005, p. 105 f.). From here follows what can be called a semantic naturalism, according to which all concepts that appear in philosophy should be reducible to scientific concepts in order to be meaningful. This is a general stance that I will adopt throughout this work and try to realise as far as possible.

The naturalistic approach to causation has become well known as the empirical analysis of causation, and the first who explicitly addressed the naturalistic method are Mackie (1980) and in more detail Dowe (2000). However, Dowe's explications have been criticised and developed further by Bontly (2006). The kind of empirical analysis that Bontly suggests starts with a Lewis-style reduction of the concept 'causation' "upon our best psychological theory of causal judgement" Bontly (2006, p. 195). After that is done, we can see whether there is any realiser for the folk theory.

In doing this reduction, one of the most important things to be kept in mind is that causation is reduced to terms that are better understood than 'causation'. For example, Bontly (2006, p. 194 f.) argues against Menzies (1996) that his reduction of causation to chance is not helpful, since chance is an equally difficult concept.

Kutach (2010) supports the same view of what an empirical analysis is, with some slight differences. Again, his aim is to identify something empirically found in the world as causation, but his starting point is the analysis of our intuitions:

> It ought to be uncontroversial that some kind of conceptual analysis is a necessary component of any intellectual investigation, for without it, we would have no way to connect our theoretical terms to the folk terms they are intended to improve and the phenomena they are intended to describe. One needs to have a requirement in one's standards for theoretical adequacy so that, for example, a successful theory of planetary motion is not passed off as a successful theory of causation merely by attaching the label 'cause' to what we ordinarily think of as an orbit. (Kutach 2010, p. 4)

Kutach sees the difference between his approach and more traditional methodologies, which he summons under the term 'orthodox analysis', in that for him not all intuitions have to come out as strictly true. Empirical analysis according to him does not search for a priori definitions, but can be content with a more or less complete collection of platitudes of which some can only be approximately realised and others dismissed completely, if, e.g., incoherent with more sacrosanct platitudes (cf. Kutach 2010, p. 5).[4]

[4] Collins et al. (2004, p. 38) come to a similar conclusion, though for different reasons. They argue that due to the complexity of the topic, to find an account of causation that coherently combines several, but not all, our intuitions is already an achievement and metaphysically interesting.

Another characteristic of Kutach’s empirical analysis is that the second part can fire back on the prior conceptual analysis. Empirical facts might serve to refine the conceptual analysis. Even though the investigation starts with a conceptual analysis, it does not follow that the latter has to be complete before the empirical investigation or that results cannot be adjusted to what has been found empirically.

> Part of what it takes to discover whether there are Ks and what they are like is determining what it would take for us to regard something as a K. This latter is a job for the conceptual analyst. [...] This does not mean that we must completely settle the question of what we would regard as a K before we write the surveys and perform the experiments. The two complementary projects—a priori conceptual analysis and a posteriori empirical examination—might interlock and cooccur. If conceptual analysis has any priority over empirical investigation, it is logical or methodological, not temporal. (Kingsbury & McKeown-Green 2009, p. 168)

As a consequence, it might legitimately happen that physics changes the way we think about causation.

To conclude this section, the question I now want to discuss is which methodology should be followed, the Canberra plan or the empirical analysis. As the previous exposition has shown, at the face of it, these two programs are incompatible with each other; while the Canberra plan starts off with a conceptual analysis presumably giving rise to a priori truths, the empirical analysis emphasises the fallible results of the natural sciences, and in accordance with its naturalistic background relegates all apriorisms. However, despite this apparent contradiction, which at the surface could hardly be greater, ambiguities in the terms ‘a priori’ and ‘naturalism’ make it necessary to take a closer look. What counts in the end are not the names, but rather how the methodologies are spelled out in detail. If attention is paid to the latter then, so I want to suggest, the border between the two methodologies becomes significantly blurry, if it persists at all.

Turning towards the Canberra plan, it might be the case that traditional conceptual analysis sees its work as being a priori in a strict sense of independent or immune to all empirical facts. However, this is not how it is understood by Canberra planners, most prominently Frank Jackson. Jackson’s ‘a priori’ is an extremely weak one and far away from how it was understood originally by Kant. According to Jackson, the meaning of concepts can on the one hand be formed historically or causally and on the other hand be revised on empirical grounds. If then, following Jackson, we can know the meaning of a concept a priori, this ‘a priori’ is dependent on the empirical world in two ways. Indeed, Lewis and Jackson understand the first step of the Canberra plan as finding our folk theory of a concept, e.g., by asking how a person in a certain situation would respond to the question whether something is causal or not. Finding this folk theory is most likely an empirical project and far from an a priori analysis.

In the light of this, how does the contradiction between the Canberra plan and the empirical analysis look? Given Jackson’s weak notion of the a priori there seems to be no difference left between how Jackson understands his project and how Bontly or

Kutach understand their empirical analysis. Jackson's understanding of it is exactly that of an a priori that as Caro & Macarthur (2004, p. 7 f.) observe, is compatible with naturalism. This is also concluded by Papineau (2009):

> The Canberra strategy thus seems no different from the prescription that philosophy should start with the synthetic theories endorsed by everyday thought, and then look to our more fundamental theories of reality to see what, if anything, makes these everyday theories true. This seems entirely in accord with methodological naturalism—philosophy is in the business of assessing and developing empirical theories of the world.[5]

In conclusion, at this point I do not see any necessity to decide between the Canberra plan or empirical analysis, since their method of investigation is in essence identical. Questions concerning semantics and, following on that, whether concepts can be analysed a priori do not matter here. Rather, what matters is that in the investigation of causation we first have to find platitudes of causation and then discuss whether these platitudes are realised, in the process of which it might turn out that this is only the case for some of them. In the present case this means that the aim is to express the rather ambiguous metaphysical term 'causation' through the well understood and empirically justified terms of physics.[6]

3.3 Platitudes of causation

The role of intuitions in philosophy is a vast topic that I cannot hope to even come close to present here with the detail that it deserves. However, as far as I can see the use of intuitions in the Canberra plan and naturalism is fairly unproblematic and not affected by the controversy about intuitions in current philosophy. Most of the discussion on intuitions centres around whether having an intuition about P justifies the belief that P is true. However, arguably the "kind of epistemic justification associated with intuitive judgments remains something of a mystery and we do not have widely accepted models of it." (Chalmers 2013, p. 10)[7] Besides this relatively problematic usage of intuitions, there is another one that is far more modest and which I will apply here, namely to take intuitions as a guide to conceptual knowledge (cf. Hintikka 1999, p. 143). Thus, intuitions about causation will not be regarded as

[5] To be fair, Papineau discusses more possibilities for dispute than just the notion of 'a priori'. These, however, rest on certain ontological interpretations of Ramsey or Carnap sentences. I do not pursue this further here, as I do not see any side in the debate necessarily committing themselves to a particular view on this topic.

[6] However, a question that will be left unanswered in this work is how far this reduction goes, viz., it constitutes a reduction of 'causation' onto non-causal concepts. The problem is that for now it is not clear whether physical concepts are non-causal. For example, it could be the case, as Carroll (2009, sec. 4.2) explains, that properties in physics are best explicated as dispositions, and thus a reduction of 'causation' onto physical properties would not be a reduction onto concepts that have no relation to causation.

[7] See Williamson (2007, p. 215) for the same position. See, however, Pust (2012, ch. 3.5) for a defence of using intuitions as evidence, and Chudnoff (2011) for the more radical position that having an intuition that P automatically justifies your belief that P.

telling us what causation in the world is, but rather they will be used as a guide to what we understand with 'causation', that is, to identify the subject matter. The methodology I am following clearly assigns only the modest role to intuitions. They only play a prominent part in the first step, which is conceptual analysis, while the task to determine whether anything in the world corresponds to the intuitions is a completely different task in which intuitions cannot lead the way, and for which one has to study physics. Thus, saying for example that intuitively causation is a chance raising relation does not mean that there must exist something like a chance raising relation in the world, but only that this is how we understand the concept of causation.[8]

I will call widely shared intuitions platitudes, and the conjunction of those platitudes is the folk theory. The problem now is to find intuitions about causation and to determine which of them should be part of the folk theory. A fairly obvious way to do so is to just ask the folk whether, given certain situations, they would describe them as causal and why. However, given the limits of this work, this option is not available. Rather, I want to rely on the intuitions of a special kind of folk that fortunately has a natural drive to express its intuitions, namely philosophers.[9] One might object to this approach that what philosophers believe about causation is so diverse that it cannot possibly lead to a coherent theory of causation. However, first of all there seems to be a large agreement on the most basic meaning of causation. This is hardly surprising, since if there was no such agreement, then philosophers arguing on the basis of different understandings of the concept would not talk about the same subject and there could be no fruitful discussion at all. On top of that, given the presented methodology, there is no obligation to capture all intuitions, but only the most common ones. The purpose of the first step in the Canberra plan and the conceptual part of the empirical analysis is to define the subject matter well enough be able to ask whether something in the world corresponds to a concept and to have a sensible discussion about that subject. It is fully compatible with this approach, and probably the case in most philosophical debates, that despite agreement on a core understanding of a concept, there is some disagreement. As I want to maintain, this disagreement happens mostly on the periphery of a concept and not on its core meaning; which is just to say that trivially the core is what is largely agreed upon. However, the point is that there is a core substantial enough to make an investigation into the concept interesting, or else there could be no debate at all. Additionally, it is methodologically admissible to decide only after the folk theory has been found, whether all platitudes of the folk theory are realised. Thus, when nothing in the world can be found that corresponds to a certain platitude about concept P, the conclusion is not necessarily that P as a whole does not apply, but only that this particular platitude of P does not.

[8]Thus, despite criticising them in the previous section, I can fully agree with Ladyman & Ross (2007, p. 12) when they proclaim that "as naturalists we are not concerned with preserving intuitions at all".

[9]A short survey of what might be the theory of the, mostly philosophically uneducated, folk can be found in Ladyman & Ross (2007, p. 268). See also Sperber et al. (1995).

Finally, it is clear that the concept of causation most likely has changed historically such that the folk theory of causation, say, 200 years ago has been different from the folk theory now. The interest of this work, however, lies in the present folk theory and thus historic accounts of causation are not considered. For example, I take it as given that causation is not necessarily deterministic, but can also be probabilistic; a relatively recent change in the concept that nevertheless is now widely accepted. It would be an interesting project to analyse in how far today's folk theory of causation has already been influenced by quantum physics. Thus, for example, whether the wide acceptance of probabilistic causation is a direct consequence of the development of quantum mechanics or whether it has been there before as, say, a consequence of the folk's view on what happens when you throw a dice.

A worry that might result from this consideration is that if today's folk theory of causation was formed by modern physics, then reducing the former to the latter involves a certain circularity. However, I do not see this circularity as damaging. What I am investigating is whether there is a correspondence between the folk theory and certain traits of physics. I am not making an argument in order to establish a conclusion on the basis of premises. If it was the case that quantum physics has generated the folk theory, then the correspondence between them would be trivial, but still there would be a correspondence and given that it is not obvious whether this correspondence exists, investigating it is a worthwhile project.

In the following sections, I will gather what I believe to be the most basic and common intuitions about causation. I am not the first one to do that, and my collection will largely coincide with earlier ones, such as can be found in Menzies (1996). The conjunction of these platitudes will be the folk theory that subsequently will be tested on whether it refers to something physical. The following platitudes are not meant to deliver a complete theory of causation, but only guiding principles. It is thus clear that many questions that often are related to causation, such as the exact nature of the relata, will remain unanswered by the platitudes. Furthermore, it is not surprising that the discussion of the platitudes is not always limited to just one platitude but sometimes interweaves different issues. Thus, for example, the topic of absences already shows up in sec. 3.3.1, but will only be treated with more care in sec. 3.3.2. To keep the discussion simple, I will nevertheless try to separate issues as well as I can.

3.3.1 Relata and relation

Relata

Following Menzies (1996, p. 98) I want to suggest that one of the basic platitudes about causation is that *causation is a relation between distinct events.* As far as I can see, this statement is shared by everyone, and denying it would amount to a change of subject, away from causation. To my knowledge, literally no one denies that causation is a relation, of whatever kind,[10] and that the relata of the relation must be distinct. Turning towards the relata for the moment, the term 'event', however, seems not that uncontroversial and needs to be qualified. In general, events can be characterised as being immanent, that is occurring in space-time, as opposed to transcendent facts (cf. Schaffer 2013). Excluding facts from being causal relata is of course disputable and philosophers such as Mellor (1987, 1995) have argued in favour of facts. However, given the present methodological background, I hold this step to be unavoidable. Since it is the goal of the method of this investigation to find something existing in the world that is causation, facts are excluded, simply because they do not exist in the world. At the same time, this does not mean that causation as a relation between facts is ruled out as impossible. It might still turn out that the method followed here does not lead to a satisfactory result, because nothing in science can be found that would correspond to the folk theory. In that case, one might be inclined to look for causation elsewhere, and that might well be among facts.

Furthermore, by saying that the relata are events, I do not wish to imply any more specified ontology. An occurring event consists of whatever entities exist, while the entities may be objects or structure or what not. As Menzies (2009) explains, events "are parasitic on physical objects in the sense that they would not exist if there were no physical objects." An event that is the relatum of a causal relation can prima facie be arbitrarily complex and consist out of multiple objects or any large structure. It is common that "several causes may be lumped into one big cause. Or one cause may be divisible into parts. [This] multiplicity of causes [is] obscured when we speak, as we sometimes do, of *the* cause of something." (Lewis 1986a, p. 215) The composition of an event out of physical things is metaphysically cheap in that it does not involve anything else than the physical things. Saying that some objects or structures form an event merely presupposes that there is some kind of distinction, e.g. spatial or temporal, between these particular objects or structures and other things that exist. It does not require that there is some novel composition relation apart from the natural relations the objects stand in or the structure involves anyway. Below I will propose that one way how the term event can be cashed out in physics is by the notion of a physical state.

Leaving the ontology undetermined is first of all a result of the fact that the platitudes of causation do not determine the ontology. Philosophers can disagree about what kind of entities are the relata of causation, or in other words, at least some philosophers promoting different ontologies share the belief in the same platitudes of causation. Also, intuitively a causal relation can hold between all kinds of things. It seems therefore not to be a part of our intuitions about causation what the relata exactly are. This does not pose a problem, since what matters is to be able to identify a relation as causal, and it is in principle possible to do this without knowing

[10]Even though some might say that certain views on causation by absence imply nonrelationalist theories of causation (cf. Beebee 2004a, p. 292). I believe this is a misnomer by Beebee and will have more to say on causation by absence in sec. 3.3.2.

exactly what kind of entities it holds between. The reason is that the identity criteria for a relation are not necessarily dependent on the identity criteria of its relata. For example, it is possible to grasp the meaning of 'being spatially separated from' without knowing what is spatially separated from what. A further example, where use of this fact about relations is made, is ontic structural realism. For example, French & Ladyman (2003, p. 39 f.) claim that what they "begin with are the laws which express the relations in terms of which the 'entities' are constituted." Thus, according to them the relation is primary, and only through the relation we get knowledge of entities.

Related to this is the question when and why events are sufficiently distinct from one another to be connected causally. The notion of distinctness has to capture the intuition that the cause is different from the effect, such that, for example, one event cannot be the cause of itself. As I will argue now, distinctness, suitably defined for causation, has two necessary conditions that together are sufficient. The first condition is distinguishability, which gives the answer to the question how we can distinguish one thing from another. As French & Krause (2006, p. 16) explain, there are two possibilities for being distinguishable, namely either if the things are different in at least one property, or if they differ in spatiotemporal location. It is worth emphasising that this is an epistemic notion, concerned with how we can know that things are distinct, rather than being about the identity of things (cf. Gracia 1988, p. 21). Distinguishability alone, however, is not sufficient, since there can be things that are distinguishable in one or two of the above mentioned aspects, but cannot stand in a causal relation. For example an atom is distinct from a positron it has as a part, both in properties and in spatial location, but we would not say that both could stand in a causal relation to one another.

In consequence, there has to be a second necessary condition, which has to point out dependence relations in which events can stand that preclude them from being causally related. According to Menzies (1996, p. 98) there are four such relations: 1. If "an event depends on a more specific event that implies it", 2. if one event is reducible to another, 3. if one event is a proper part of another, and 4. "when an extrinsically characterised event depends on an intrinsically characterised one". Thus, for example, the event that Peter is a man cannot cause him to be a bachelor (1), a shadow cannot cause there being light of lower intensity (2), the event that there is a marriage cannot cause there being a groom (3) and that the Eiffel tower is higher than the Dome of Cologne cannot cause it to be 324m high (4). In his (1988), Menzies argues that these kinds of dependence relations are supervenience relations. Even though I do not disagree with Menzies, I want to suggest that they are best subsumed under a different more general header, namely that of existential dependence. According to Lowe (2010) a thing, e.g., an entity or an event, is rigidly existential dependent on another if the condition (EDR) holds:

> (EDR) x depends$_{\mathrm{R}}$ for its existence upon $y =_{\mathrm{df}}$ Necessarily, x exists only if y exists.

(EDR) includes all three cases of dependence that Menzies identified, but it is also more general, since it is a symmetric relation.[11] I do not see, though, that this generality would exclude any cases that one would intuitively name causal. Therefore, the second necessary condition for distinctness is that cause and effect are existentially independent of one another. One might object that in general existential independence implies distinguishability and thus the former condition alone is sufficient. However, this is not necessarily the case, as for example it has been argued that in quantum mechanics there are existentially independent particles that are at the same time indistinguishable. In conclusion then, distinguishability together with existential independence is a sufficient condition for the relata of causation to be distinct.[12]

A possible way in which the above explicated notion of events can be cashed out in physical terms is by referring to the relata of causation as states. According to Chisholm (1992, p. 4) what a state is can explained by the following statement:

> For every x, if x exemplifies being-F, then there is a state, x-being-F.

Where x can be any entity and F any predicate. Chisholm further explicates that things "may be said to enter into temporal and causal relations *via* their states." (Chisholm 1992, p. 4) Additionally, I want to stress that, following Chisholm (1992, p. 5), the notion of a state is different from that of an individual thing. A state does not provide the identity criteria necessary to fully categorise what a state is ontologically. Given that identity conditions sort entities into more determinate kinds, a state is indeterminate with respect to its ontological kind (cf. Lowe 2006, p. 6). For example, a state can be that of an object, or the state of a structure. The concept of a state seems to be easily applicable in physics, being a set of physical properties (cf. French & Krause 2006, p. 17). However, how exactly states in physics can be distinguished and whether they are useful as causal relata in QFT is a matter that needs to be further discussed (see ch. 5).

Relation

Having hopefully explicated with enough detail to what the first intuition about causation amounts to for the relata, I will now turn to the relation. Again I agree with Menzies that one of the central platitudes about the relation is that "the causal relation is an intrinsic relation between events." Menzies (1996, p. 98)[13] According

[11] This leads Lowe (2010) to conclude that it is unsuitable to capture our intuitive notion of existential dependence. However, this does not need to concern us here, since the task is not to define dependence, but rather distinctness suitable for causation.

[12] See also Lewis (1986c, p. 258), who comes to the conclusion that two events are not distinct, and thus one cannot be the cause of the other, if one is part of the other or if they share a part. This includes logical implication as well as proper parthood. To take examples from Lewis, if someone writes "Larry", then writing "Larry" cannot cause writing "rr", nor can writing "Larr" cause that person writing "rry". Also according to Lewis, molecules are not distinct from the atoms they are made up from.

[13] Indeed, for Menzies this is a different platitude from the one that causation is a relation between distinct events. However, I believe that nothing depends how exactly the platitudes are ordered, as long as all important intuitions are captured somehow.

to Menzies, this means that the relation supervenes on the natural properties of the relata and the natural relations holding between them. Thus, a particular causal relation depends only on the events that constitute the cause and the effect and not on any other events. This makes it unnecessary to consider facts about, e.g., preemption in which more than two events are involved. According to Menzies it is simply a fact of the folk's intuition of causation that the latter is intrinsic, and that it therefore contradicts what are possibly the most popular theories of causation in philosophy, namely, counterfactual theories.[14] Causation of the intrinsic kind has been called causation as production by Hall (2004b, p. 225), a term that obviously needs to be clarified further at some point, as opposed to the extrinsic counterfactual dependence. Even though Hall (2004b, p. 226) admits that our intuitions concerning counterfactual dependence are not very firm, he argues contrary to Menzies, who dismisses counterfactual dependence for the above reason, that there are two kinds of causation, production and counterfactual dependence and that we cannot do without the latter. This is for two reasons, first, we need counterfactual dependence to describe causation by absence, and second, causation as counterfactual dependence appears in causal decision theory. Absences will be the subject of section 3.3.2 where I will argue that they do not necessarily lead to a counterfactual theory of causation.

So what about decision theory, does it require counterfactual causation? According to Hall we often use counterfactuals to estimate what would be the outcome of certain actions, and decide to act accordingly, so that the most desired outcome happens.

> When you face a range of options, and causal decision theory says (very roughly) that the rationally preferable one is the one most likely to have as a causal consequence the best (by your lights) outcome, the notion of 'causal consequence' at work is clearly that of dependence, and not production. (Hall 2004b, p. 268)

Hall's example is that of Suzy who is the pilot of a bomber, en route to destroy a target. Far away Billy, in his fighter plane, interrupts an enemy plane that otherwise would have shot down Suzy and thus would have prevented the bombing. Billy prevents the enemy from preventing the bombing; a case of double prevention. This leads Hall to the following conclusion:

> Our standard stories of double prevention already illustrate the irrelevance of production to decision making. There is, we can suppose, nothing whatsoever that Billy can do to help produce the bombing, but that doesn't matter in the slightest: Whether there is a bombing clearly depends on the action he takes, and it is his beliefs about this dependence and its detailed structure that will guide his decisions, insofar as he is rational. (Hall 2004b, p. 269)

The first comment I have to make is that just because something is called a 'causal consequence' it does not need to be causal in a relevant sense. Be that as it may,

[14]Building on this, one can argue against counterfactual theories of causation on the basis of that they contradict too many or too crucial platitudes of causation to be called a theory of causation, because they are not only extrinsic, but also fail to be transitive (cf. Hall 2004a for the argument on transitivity).

I want to maintain that Hall is wrong in claiming that production is irrelevant to decisions. Even though it is true that Billy cannot help to produce the bombing, it certainly matters to him that there is a possible line of causal production that would lead to the shooting down of Suzy and that he can prevent. Thus, Billy can come to a decision by comparing different lines of causal production, some of which he plays an active role in, some of which he remains inactive in, and then choosing the one with the best outcome. I do not see why it should be necessary that Billy posits some more or less direct causal connection between his actions and the bombing. Consequently, contrary to what Hall claims, coming to a decision in situations like the one of Billy, does not presuppose a theory of causation as counterfactual dependence.

On top of that, even if it were true that Billy necessarily has to assume a line of counterfactual dependence between his action and Suzy, causal decision theory would still not be enough to call this kind of dependence causation. The reason is that it is precisely the question whether counterfactual dependence is causation or not and causal decision theory gives no additional argument for that, it simply uses counterfactuals. Only when there are already independent reasons to regard counterfactuals as causal, Hall's argument works. On the other hand, even if there are convincing reasons to assume that causation is necessary to make decisions (cf. Cartwright 1979), this does not mean that the theory of causation involved has to be a counterfactual one (cf. e.g. Price 1991). In other words, Hall taking causal decision theory as an argument already presupposes that counterfactual dependence is causal and thus he makes a circular argument. As a consequence, I will side with Menzies and regard causation as an intrinsic relation.[15]

The final platitude on the causal relation, which I briefly want to mention, is that of transitivity. When I bump into the table which causes the coffee to spill which then makes the book wet, then there is a strong intuition that me bumping into the table is the cause of the book being wet. If a causes b and b causes c then a is the cause of c. (cf. Hall 2004b, p. 228) As Hall notices, transitivity can easily be turned into an argument against theories of causation, if not all relevant causes for an event are taken into account. For example, the ship's radar system enables it to steer safely through the traffic in the English Channel despite the fog, but later in the Arctic the first officer falls asleep, doesn't watch the system and the ship hits an iceberg. Thus, counterintuitively, the ships radar system is the cause for its hitting the iceberg, because it prevented it from colliding with another ship in the Channel earlier, which would have prevented the ship from reaching the Arctic. The consequence of this is that only when 'cause' is used in the "egalitarian" (Hall 2004b, p. 228) sense, that is, all causes are subsumed, then causation is fully transitive. For the above example, 'cause' in the egalitarian sense then means that the radar system and the first officer's sleep, among other things, cause the ship to hit an iceberg, but

[15] I wish to point out that I have not presented any knock down arguments against counterfactual theories of causation. I have criticised the most common argumentations for such a theory, but it is of course still possible to stick with counterfactual causation for different reasons.

not the radar system alone.[16]

I wish to end this section by considering a possible objection. Menzies (2009) makes an argument to the effect that the initially stated platitude cannot be crucial for causation. He first observes that causation usually is not only a relation between distinct events, but also a *natural* relation. It follows that cause and effect have to be natural events, that is, they have to exist mind-independently and 'carve nature at its joints', also there should be an objective description of these events. Giving various examples, Menzies then argues that the description of the events often is not objective, but rather depends on the context and normative or other background conditions. For example, in the eyes of the peasant, the drought caused the famine, but in the eyes of the United Nations it was the government's failure to take appropriate action. Or the man who stood on the shore can be made responsible for the drowning of the child, because he should have helped, and so on. The conclusion Menzies draws from here is that the platitude that causation is a natural relation between distinct events cannot be the basis for a theory of causation. Instead, Menzies proposes to analyse causation as a relation that makes a difference from the normal course of events.

Since, Menzies' alternative is very close to traditional interventionist accounts of causation I will treat it further in sec. 3.3.4 and explain there why I do not agree that causation can be reduced to difference making. At this point I merely want to argue that the observation that causal judgements often depend on context is not a good reason to abandon the first platitude on causation. First of all, the examples Menzies relies on seem to be at odds with the premise that causation is a natural relation. If one starts out with the requirement that there is an objective description of cause and effect and that they carve nature at its joints, then examples from everyday life are ill suited to illustrate these events. The way we speak in everyday life is very often subjective and anthropomorphic in multiple ways, and it thus cannot be surprising that we do not find an objective and mind-independent description of events here. Rather, if one wants an objective description of natural events, then one should look into an objective science that describes what exits, which obviously is physics. Without going into details, I assume that examples involving normative background conditions or subjective descriptions of events do not occur in physics.

Now, this is not to say that the examples Menzies gives do not show causal relations. Instead the conclusion I want to draw is the following. Either one can

[16] An intuition that I hold to be non central, but at least worth mentioning, is that of the number of relata. I assume that in general the intuition is that there are two relata, the cause and the effect. However, e.g., Hitchcock (1996, p. 416) has claimed that causation "is not a *binary* relation between events." An overview of reasons for denying binarity can be found in Schaffer (2013, ch. 1.3). Without going further into this topic, as it would be too voluminous to be treated here, it seems that all these reasons depend on the specific theory of causation that is adopted. Thus for example, the argument Hitchcock (1996) makes depends on his theory of causation as chance raising relation. I suggest that the methodologically sound way to handle the number of relata is to assume for the time being that the number is two, leaving open the possibility that after taking sides for a certain theory of causation there might be reasons to change this position.

insist that causation is a natural relation, but then one should only look into physics for examples. Or one can just leave out the 'natural' and, if not a single theory of causation can be found that incorporates all examples one wishes to cover, admit that causation is a multi level concept and that, e.g., normative conditions only play a role on a higher level. For the moment I do not see any necessity to side with any of these possible answers and I do not claim that there cannot be more possible replies. I am merely content with the conclusion that Menzies' argument is not a compelling reason to give up the first platitude. In so far as the present work is only concerned with physics, Menzies' concerns do not seem to apply.

3.3.2 Absences

Absences, either as causes, in which case they are called omissions, or as effects, in which case they are called preventions, have been looming large at various places in the last section. For example Lewis (2004b, p. 281) has accused Menzies (1996, 1999) of failing to successfully completing the Canberra plan, because he did not include causation by absence. Hall (2004b) has argued that absences are one reason why counterfactual dependence has to be accepted as a kind of causation. And Mellor (1995) makes the point that since absences cannot be events, causation by absence is a reason to take facts as causal relata. Contrary to these philosophers, I will argue that at least in physics there is no need to include causation by absence into a theory of causation.

The arguments in favour of causation by absence all rest on the intuition that absences can play a role as causal relata. For example, Mellor (1995, p. 134) is certain that considering the facts " 'Don dies because he falls' and 'Don survives because he holds on' [...] each is as obviously causal as the other: if either of them is a causal statement, both are." Don's omission to hold on causes him to die, and in the other case Don's holding on prevents him from dying. Another prominent example of negative causation is Lewis' void: "The void is deadly. If you were cast into a void, it would cause you to die in just a few minutes." (Lewis 2004b, p. 277) The problem now is, how to incorporate absences into a theory of causation. In the terms of Lewis (2004b, p. 281):

> A relation requires relata. The void affords no causal relata: There's nothing there at all, so there's nothing for events to happen to, so the void is devoid of events. And even if we allow causal relata to belong to other categories, still there would be none of them in the void—because there's nothing at all in the void.

Obviously, the problem is particularly acute for theories that regard causation as a natural relation between events, while on the other hand it seems easier to be incorporated into counterfactual theories (cf. Lewis 2004b, p. 283; Hunt 2005). I do not deny that there is a strong intuition that absences can be causal relata, however, I will argue that at least for physical processes there is a better alternative.

To begin with, I want to discuss two examples of how negative causation could be handled, that of Dowe (2000) and Beebee (2004a), neither of which I endorse.

According to Dowe (2000, p. 125) there is what he calls an "intuition of difference" between omission and prevention on the one side and other instances of causation on the other side. As a consequence, Dowe suggests that negative causation is not real causation, but something different that he calls causation*.

> My claim is that causation* should be understood not as real causation but as a hybrid fact usually involving certain actual real causation together with certain counterfactual truths about real causation [...] In other words, I am offering a counterfactual theory of prevention and omission. (Dowe 2000, p. 124)

Thus when we say that the rain prevented the fire from destroying the house, the causal connection is not between the rain and the absence of a destroyed house. Instead, there is a causal* fact that the fire would have destroyed the house if there had been no rain, and there is the causal fact that the rain extinguished the fire.

A different strategy to handle negative causation is promoted by Beebee (2004a). Her proposal is to regard negative causation as instances of causal explanations. Given that "not all causal explanations are reports of causation" (Beebee 2004a, p. 301), it can be consistent to regard negative causation as a form of causal explanation without regarding it as a form of causation. Beebee follows Lewis in that "*to explain an event is to provide some information about its causal history.*" (Lewis 1986a, p. 217)[17] Crucial is the claim that for an explanation we merely need 'some information' about the causal history, but not necessarily the causal history itself. Following Beebee (2004a, p. 302) it is possible to "give information about an event's causal history in all sorts of other ways—by saying, for instance, that certain events or kinds of event do not figure in its causal history, or by saying that an event of such-and-such kind occurred, rather than that some particular event occurred." Thus for example, the void did not cause the astronaut's death, but saying that the void caused it is a causal explanation, since it gives information on the causal history of the death. If this is confusing, then one can make the distinction between causation and causal explanation explicit by making a difference between 'cause' and 'because'. (Beebee 2004a, p. 306) Accordingly, in the present case it would be better to say 'The astronaut died, *because* of the void', highlighting that the void merely is part of the causal explanation.

How convincing are Dowe or Beebee in the light that there is the intuition that absences can be causes and effects? Are there any reasons to believe that even though negative causation might be part of the folk theory, it is a platitude that is not crucial and does not have to be part of the final theory of causation? Dowe does not give any arguments other than insisting that there is an intuition of difference.

[17] Indeed, I see Beebee's solution already anticipated in Lewis (1986c, p. 268 f.), who writes: "If there are no extrinsic or disjunctive events to be caused, still there are extrinsic or disjunctive truths about regions to be explained. [...] Explanatory information about the explanandum truth consists in part of non-causal information about the truth-making pattern itself: what sort of pattern it is, and what events comprise it. And it consists in part of information about the causal histories of the events that comprise the pattern. As usual, explaining means providing some explanatory information."

On the other hand, one reason to exclude negative causation according to Beebee (2004a, p. 293) is that including it would lead to unintuitive results. For example, one can regard Mary, who has spend her entire life in Sydney, as a cause why Peter, who lives in New York and has never heard of Mary, forgot to go shopping yesterday, because she omitted to remind him. Going further this way, one ends up with the very unintuitive result that there is an enormous number of negative causal relations, since not only Mary failed to remind Peter, but also Bob and Paul and everyone else as well.[18] Intuitively one would like to draw a distinction between Mary and Peter's wife Rebecca, since the latter actually promised to remind him, but did not do it. However, according to Beebee (2004a, p. 300) this is not possible, since there "isn't any objective feature that some absences have and others lack in virtue of which some absences are causes and others are not." It follows that any account of causation that wants to treat causation by absence as objective must fail. Beebee concedes, though, intuitively we can make distinctions between absences, for example, on the basis of moral obligations. However, she rejects that these kind of criteria should play a role in the metaphysics of causation. For her the consequence is that there probably is no theory of causation, which always gives intuitively correct results when it comes to absences. This leads her to the conclusion that "there are features of commonsense assertions [...] to which *no* theory of the metaphysics of causation ought to be doing justice" (Beebee 2004a, p. 291).

However, Dowe's and Beebee's arguments to discard intuitions on negative causation certainly would not convince everyone. The accounts of Dowe and Beebee are similar in that both rename negative causation and then treat it differently from genuine causation. Dowe on the ground of the 'intuition of difference' and Beebee to avoid unintuitive results. Renaming negative causation is a move that is strongly rejected by Schaffer (2004, p. 197): "Negative causation is genuine causation, or so I shall argue." However, even though Schaffer promises an argument, what he actually does is merely presenting a number of examples, in which "[e]xpert psychologists, biologists, chemists, and physicists routinely use 'causes' to cover cases of negative causation." (Schaffer 2004, p. 208) Schaffer gives examples of cases that intuitively are clear cases of negative causation, and he insists that this intuition cannot be dismissed. In his eyes, Dowe and Beebee give no compelling reasons why intuitions about negative causation should be treated differently from intuitions on other cases of causation.[19] Nevertheless, even though Schaffer does not give an argument, reminding us that there is a strong intuition in favour of negative causation has some force. Prima facie, a strong intuition should be accounted for in a theory of causation. At the very least it is clear that simply renaming negative causation without a good reason is no compelling strategy, because as Schaffer can always can insist: "I only ask that the empirical philosopher not ignore intuitively paradigmatic

[18]For McGrath (2005, p. 125) this presents a dilemma: "Either there is no causation by omission, or there is far more than common sense says there is."

[19]Schaffer also argues that Dowe's proposal cannot handle more complicated cases of negative causation, such as involving overdetermination. Since I do not endorse Dowe's view on negative causation, I do not discuss this critique further.

cases of negative causation". (Schaffer 2004, p. 208) The same objection comes from Menzies (2009), who points out "that there is no psychological evidence that people draw a sharp distinction between causation involving positive and causation involving negative events."

In summary, there is a standoff between Dowe and Beebee on the one side, and Schaffer on the other. Dowe thinks that there is an intuition of difference. Schaffer denies that there is such an intuition. Beebee thinks that accepting negative causation leads to absurd consequence. Schaffer thinks that denying negative causation is absurd. In what follows, I will argue for a way out of this situation.

Is there any way in which Schaffer's claim that negative causation is genuine causation can be refuted? Schaffer gives no argument, but only insists on the intuition that there are clear cases of negative causation that are on par with positive causation. There are thus no premises to attack and simply denying the intuition will not convince anyone. In my eyes, the only way to argue against Schaffer's claim is to have his intuition trumped by another intuition. Accordingly, I will argue that at least in the cases that matter here, namely fundamental physical processes, there is another intuition that dominates over the intuition that absences can be causal relata.

Let us take a closer look on one of the prime examples again, Lewis' void. The void is deadly, but Lewis realises that what happens to a person in the void can be further explained with the help of physics.

> What I've said is literally true, yet it may be misleading. When the void sucks away the air, it does not exert an attractive force on the air. It is not like a magnet sucking up iron filings. Rather, the air molecules collide and exert repulsive forces on one another; these forces constitute a pressure that, if unresisted, causes the air to expand and disperse; the void exerts no force to resist the pressure; and that is why the air departs from the lungs. (Lewis 2004b, p. 277)

The other effects that being in the void have on the body are explicated in a similar fashion by Lewis. I believe that Lewis goes in the right direction by admitting that one can be physically more precise, but he does not go far enough. He still explains what happens as an omission by the void, e.g., not to exert pressure. Contrary to that, it is possible to tell a complete story about what happens solely on the basis of positive events, without mentioning the void. Thus, the air expands not because the void does not exert pressure on it, but only because of the movement of the air molecules. The movement of the air molecules is the cause and expansion the effect. Likewise, the blood does not boil, as Lewis would have it, because "the void exerts no counterpressure" (Lewis 2004b, p. 277), but because in the given condition blood has a lower enthalpy of vaporisation and molecules in the blood have enough kinetic energy to escape from the surface of the blood. Another example can be found in Schaffer (2004), who claims that electron-hole pair creation is an example of negative causation in physics. However, as Ney (2009, p. 760) observes, this can easily be translated into claims about positive charges. As I want to maintain, this is the case for all seemingly instances of negative causation in physics; all can be translated into stories of positive causation. Absences simply do not occur in how fundamental physics describes the world.

As Schaffer (2004, p. 210) observes, "intuitions come in degrees: some intuitions are paradigmatic near-certainties; others are mere borderline tugs." At least in physics, the intuition that absences can be causes or effects, is far shakier and borderline as Schaffer wants to have it, and I want to maintain that there are other intuitions trumping them. The most important intuition is put into the short form "Nothing will come from nothing" by Armstrong (2004, p. 48). Even though intuitively we might count cases of omission and prevention as cases of causation, there is also an intuitive reluctance to admit that nothingness, like the void, can have any effects. If one had to choose, I suppose that no one would entirely dismiss physically real events entirely in favour of absences. Accepting absences as well as physically real causes, however, leads to a kind of causal overdetermination. For example, the molecule's kinetic energy, as well as the void would be the cause of the vaporisation. Causal overdetermination is something that always should be avoided and therefore a choice has to be made. I think that taking physical real events, like a molecule having a momentum, as the cause or effect is far more intuitive than taking the void.

In opposition to this, Schaffer (2004, p. 210) argues that going from 'the void causes' to 'the thermal energy of blood causes' is begging the question, since it "implicitly assumes the physical connection view". This, however, is not the case. The move of translating a negative event into a positive one only acknowledges that there is a different description of what happens, and that only the physically accurate one is explicit about which entities exist.[20] Whether this different description should be treated within a process or a counterfactual theory of causation, or is not causal at all, is a different matter; both would be possible.

In conclusion, I think that every case of negative causation in fundamental physics can be translated into a case that does not involve absences, and that the latter is far more intuitive. It is noteworthy that, nonetheless, this understanding of causation by absence can be conjoined with theories that explain why we sometimes talk as if absences were causes or effects. For example by employing Beebee's (2004a) view that talk about absences give explanatory relevant information or, as Menzies (2009) argues, that absences give contrastive information. What matters most at this point is that even though we have the intuition that there is negative causation, a theory of causation in fundamental physics does not have to count absences as causal relata.[21]

[20] My solution is similar to what Hunt (2005, p. 218) proposes, drawing on Davidson's work on events: "This dilemma dissolves once we distinguish between the event and our description of it as an omission. By describing what the father did as omitting to pay attention to his child we imply that the father was either tying his shoelaces, intently gazing into the sky, or doing anything that similarly precludes paying due attention to his child. The description may be implicitly disjunctive but the actual event is, say, tying his shoelaces, which is not 'disjunctive' at all." Though Hunt sees his way out as more general and not only applicable in physics.

[21] I belief that this proposal fits nicely into a theory of truthmaking according to which a proposition can be true, even though parts of that proposition have no referent. A guiding example would be that of French (2014, sec. 7.4.2.3), who argues that propositions about objects can be true,

3.3.3 Probability raising

Intuitively it seems obvious that causes raise the probability for their effects. El Niño increases the risk for a drought in certain parts of Australia; speeding raises the probability of an accident, and so on. Furthermore, the rise of quantum physics has convinced many that the world is not deterministic and thus stirred new interest in indeterministic causation among philosophers. This induces the question whether causation can be reduced completely to a statement about probabilities. In particular, the question is whether there is a probabilistic theory of causation not only for general causation[22] that handles probabilities in terms of relative frequencies, but also for singular causation. I do not think that there is such a theory and will outline the reasons in this section. It does not follow, however, that probabilities would be irrelevant for causation. Causation cannot be reduced to probabilities, but the intuition that causes raise probabilities cannot be discarded. Thus, my main concern is to explicate in how far a theory of causation should agree with the initial intuition, without being reductionistic.

As an obvious starting point, the central intuition can be captured by the following reductive statement:

C causes E iff $\mathrm{P(E|C)>P(E|\neg C)}$

That is, C causes E iff the probability for E given that C is higher compared to given that not C.[23] However, as already noted by Reichenbach, this condition cannot be the whole story, since there are clearly cases in which two events are positively correlated, but neither causes the other. For example, water in a kettle is more likely to be bubbling when it is steaming, but steaming does not cause the bubbling. Rather, that the kettle is heated up causes both the bubbling and the steaming. Reichenbach (1956, ch. 19) therefore introduced the principle of the common cause: "*If an improbable coincidence has occurred, there must exist a common cause.*" Reichenbach (1956, p. 157) More formally, a common cause structure has to fulfil the following conditions:

$$\mathrm{P(A.B)>P(A)P(B)}$$

$$\mathrm{P(A.B|C)=P(A|C)P(B|C)}$$

According to the first condition, the probability that events A and B occur together is higher than what could be expected, if they were independent of one another. The second condition tells us that the probability for A and B to occur together if C occurs is as high as the one in the case where A and B are completely independent

even though there are only structures. Analogously, propositions about absences can be true, even though absences do not exist.

[22]General in the sense that multiple instances of a relation have to be considered in order to make an inference about the causal content of a single instance.

[23]For details and early versions of probabilistic theories of causation see Reichenbach (1959); Good (1961); Suppes (1970) and the survey and critique in Salmon (1980b).

of one another and each occurs together with C. For this reason C is said to screen off A from B.

As Cartwright (1979) has pointed out, it often happens that causes do not raise the probability for their effect, because they are correlated with another factor that lowers the probability for the same effect. These cases are known under the name of Simpson's paradox. To take Cartwright's example, smoking increases the risk of heart disease. If, however, it were the case that smokers usually do a lot of sport, then it could turn out that in total the risk for heart disease is lowered by smoking. The latter is the case if doing sport is more effective in preventing heart disease than smoking is for causing it. Cartwright proposes that counterintuitive situations like these can be avoided, if causal claims are tested only in populations that are homogenous with respect to all causally relevant factors, except the one that is tested. Thus, if one wants to test whether smoking causes heart disease, then one has to chose either a population in which everyone exercises or a population in which no one does, and only vary whether people are smoking. As a consequence, Cartwright's probabilistic theory of causation is the following:

> *C causes E* if and only if C increases the probability of E in every situation which is otherwise causally homogeneous with respect to E. (Cartwright 1979, p. 423)

Even though this solves Simpson's paradox, Cartwright has recognised that this definition cannot serve as a base to which causation can be reduced. This is because the specification of what a causally homogeneous situation is can only be done in causal terms, thus rendering the whole theory circular (cf. Cartwright 1979, p. 423 f.).[24]

On top of that, the discussion has shown that there are other examples of probability lowering causation that at present cannot be accounted for in any reductive probabilistic theory of causation. Consider the following example, found in Dowe (2004): There are two ways in which atom a can decay into atom d, either via first decaying into atom b or via first decaying into atom c. The probabilities for each decay to happen within a certain time interval are:

$\mathrm{P}(a\text{-}b)=\frac{1}{2}$
$\mathrm{P}(a\text{-}c)=\frac{1}{2}$
$\mathrm{P}(b\text{-}d)=\frac{1}{4}$
$\mathrm{P}(c\text{-}d)=1$

[24] Cartwright's solution to Simpson's paradox is obviously a general theory of causation in the sense that it is not necessarily applicable for singular cases, as opposed to the first proposal in this section. Thus, even if Cartwright's solution would be feasible, it would not automatically solve Simpson's paradox for singular causation. A probabilistic theory of singular causation that can handle Simpson's paradox can be found in Lewis (1986b, Postscript B). Lewis analyses causation in terms of probabilistic counterfactuals, the truth value of which is evaluated using possible worlds. For the sake of brevity, I will not discuss this theory further, but simply assume that Menzies (1989) is correct in arguing that it falls prey to the same problems that I present at the end of this section.

It follows that the probability for the decay a-b-d is $1/8$ whereas the probability for a-c-d is $1/2$. Thus the decay a-b lowers the probability for d to occur in a given time interval. According to Dowe (2004, p. 34), the general structure of such cases is that there are two possible ways how the cause can lead to the effect and the process that actually happens prevents a more probable process from happening. In this case, the process a-b as well as leading to d prevents the decay a-c from happening, which would have lead to d with a higher probability.

This motivates a possible argument to the effect that there are two causal lines, one of actual causation and one of preemption, and that somehow only the actual causal route is relevant; turning a case of probability lowering causation into a case of probability raising causation. In the decay case this means that since only the decay a-b actually happens, the probability of a-c does not have to be accounted for, and as a result the probability for d in the case of a-b is at least higher than in the case of not a-b.

It is, however, not clear whether a move like this is compatible with a theory that reduces causation to a probability raising relation, since the distinction between causal lines cannot easily be done without invoking more than a probabilistic theory of causation.[25] Nevertheless, there is a variety of other proposals that have been made to solve the problem of probability lowering causation. Salmon (1998, ch. 14) has laid out three different solutions, the first is "the method of more detailed specification of events" (either the cause or the effect), the second is "the method of interpolated causal links", and the third is the method of "successive reconditionalisation". However, replying to Salmon, Dowe (2000, 2004) has shown that none of these solutions works in the present case of decaying atoms. First, it is not possible to further fine-grain the cause or the effect, since we are already at the finest level in physics, and second, there exist no intermediate events that could be interpolated or reconditionalised on. Furthermore, there seems to be no general rule as to how fragile events have to be, which makes Salmon's solution ad hoc, if the fragility is chosen just so the probabilistic theory of causation turns out to be correct.

Another possible solution can be found in Hitchcock (2004a, p. 405), according to which, similar to Dowe's analysis, claims about probabilistic causation always contrast an actual causal relation with a class of contrastive possible cases in which the actual causal relation does not take place. This contrastive class can be chosen such that the probability for the actual case is either higher or lower compared to members of the contrastive class. As Hitchcock argues, there is no objective way to determine the class, and thus there is not objective fact whether an actual causal relation is probability lowering or raising.

Yet another line of argument against probability lowering causation is what Dowe (2000, p. 38) calls the despite defence. For example, Eells (1991, p. 295) argues that whenever an event a lowers the probability for an event b, then b does not happen as an effect of a, but despite a. Therefore, probability lowering relations

[25] See Hitchcock (2004b) for a positive, but Dowe (2004) for a negative answer to this question.

are really not causation, but a different kind of relation. Similar to that, Beebee (2004b) argues that the relation of probability lowering causation is not causation, but a different kind of causal relation, called hindrance. Dowe has argued against the despite defence on the ground that "'despite' does not unambiguously denote a negative cause" (Dowe 2000, p. 39) and that, to return to the decay example, intuitively it seems to be unwarranted to say that atom a can cause d via c, but not via b, which would be necessary for maintaining that a decays to d despite of b. Furthermore, it can be argued that the despite defence would turn causation into a non-transitive relation, for example, if A causes B, C happens despite B and C causes D, then A does not cause D.[26] Be that as it may, I do not discuss these ways to defend probabilistic theories of causation further, since I believe that the following examples show that probability lowering causation is not at the heart of the problem for probabilistic theories in general.

Consider the following experiment (a variation on what can be found in Hitchcock 2004a): A piece of DNA has a certain probability to mutate when exposed to radiation. In an experiment the DNA is set close to two atoms, one is uranium and one is plutonium. Both have a different probability to decay and emit radiation sufficient for the DNA to mutate. As it happens, within a given time interval only the uranium atom decays and the DNA mutates. Now, following a standard probabilistic analysis, the presence of the plutonium atom actually raised the probability for the DNA to mutate, since the chance for the mutation in the presence of both atoms is higher compared to the presence of only one. Intuitively, however, it seems odd to say that the plutonium atom caused the mutation, given that only the uranium actually decayed. What we have is a case in which a potential cause raises the probability for an effect without causing it, and it shows that probability lowering causation is not the only problem for a reductive probabilistic theory of causation. However, this case of decaying atoms affecting the DNA is similar to the previous one of probability lowering causation in that preemption again plays a role in it. The plutonium atom does not cause the DNA to mutate, because it is preempted by the uranium atom.

This suggests that the problem can be solved by arguing against preemption, e.g., by insisting on that causation is an intrinsic relation (as I did in section 3.3.1 above). However, as the following example shows, preemption is not always involved. Consider a case similar to the previous one, only that now the plutonium and the uranium atoms not only have a probability to decay within a certain time interval, but also to emit radiation only in one particular direction. As it happens, out of indeterministic quantum mechanical processes, both atoms emit radiation at the same time, but only the uranium atom emits in the direction of the DNA, causing the latter to mutate. In this case again the presence of the plutonium atom raised the probability for the mutation without causing it, but now the uranium atom did not preempt the plutonium.[27] The upshot of this example is that neither preemption

[26] A reply to Dowe's charges can be found in Beebee (2004b).

[27] This example is of course similar to the often discussed scenario of two gunmen shooting at a

nor chance lowering causation is at the heart of the problem, rather the analysis of causation as probability raising seems dubitable as a whole.[28]

A popular reaction to the shortcomings of reductive probabilistic theories of causation is to give up the hope to reduce causation solely to probabilities and to combine statistical relevance with a process theory of causation, as for example in Menzies (1989), Beebee (2004b) or Hitchcock (2004a). I am sympathetic to these proposals, though for now the general lesson from this discussion I wish to draw is not so much the turn towards a particular theory of causation. On the one hand, a result of this section is that "causal concepts cannot be fully explicated in terms of statistical relationships" (Salmon 1989, p. 168). On the other hand, I wish to suggest that this is not a sufficient reason to completely dismiss the intuition that causes raise the probabilities for their effects. Rather, one can say with Menzies (1996, p. 101) that causes *typically* raise the probability for their effects. Thus, statistical relevance is neither a sufficient nor a necessary condition for causation, but we can expect that it holds for most cases. It is therefore admissible that some causes do not raise probabilities and that some probability raising events are not causes.[29] This, of course, does not preclude the possibility to find a theory of causation according to which in all cases of causation the cause raises the probability for its effect, even though the theory does not reduce causation onto probabilities. Given the intuition that causes raise probabilities, this should be regarded as a virtue.

3.3.4 Manipulation and intervention

It is intuitively clear that causation is strongly connected to manipulation. One can turn the light on and off by flipping the switch; accelerate the car with the gas pedal; boil water by turning on the stove, and so on. In short, the effect can be produced or manipulated via the cause. As von Wright (1971) notices, this connection between agency and causation historically even has had an impact on language. "The way a cause operates is often compared to the operation of an agent, who is held responsible for what he does. Some historians of ideas [...] maintain that the ancient Greeks modelled their idea of causation in nature by analogy with ideas from the realm of criminal law and distributive justice. The cause brings about a disturbance of a state of equilibrium and is thus responsible for some evil or wrong in nature." (von Wright 1971, p. 64 f.) This understanding of causation might also

target, where both have only a certain probability to hit (see scenario 6 in Hitchcock 2004a).

[28] I suggest that this is due to the irreducible indeterminism that quantum physics brings with it. Whenever a given cause, like the decay of an atom, can have different effects or not happen at all, the connection between causation and statistical relevance is completely broken. It is impossible to read off from the statistics what the actual causal connections are.

I acknowledge that it is methodologically questionable to take intuitions as a guide in conceptual analysis, and then argue against an intuition on the grounds of physics. However, as argued in sec. 3.2 this is admissible to a certain extend. Furthermore, it should be noted that the point I wish to make in this section does not hinge on quantum physics and can be made also using everyday world examples, as, e.g., in Hitchcock (2004a).

[29] This does not mean that probabilistic reasoning would be insignificant and should be replaced by causal reasoning altogether. See, e.g., Gillies (2001) for a defence of probabilistic reasoning.

be found in Aristotle’s distinction between natural and violent locomotion, where the latter can happen by pulling, pushing, carrying or twirling and always requires a continuous mover (cf. Aristotle 1930, 243a). Heidelberger (1992, p. 139) highlights that in the last century, the debate whether molecules and atoms exist, centred around the physicist Jean Perrin, has brought an interventionist understanding of causation back into the focus of philosophy. One of Perrin’s central arguments was that the best explanation for observed experimental results is that experimenters can intervene on atoms and molecules, which then cause the results. Consecutively, the connection between causation and agency has been exploited by several authors (e.g. Collingwood 1940; Gasking 1955; von Wright 1971) to analyse the former by the latter. I will start this section by discussing the more recent agency theory of causation by Menzies & Price (1993) and, after dismissing it, continue with the interventionist theory by Woodward (2003). As before, the aim of this section is not to spell out a full theory of causation, but to explicate the intuitive connection of causation to agency and, if possible, turn it into a general condition that a theory of causation has to agree with.

Agency causation

The central claim of Menzies & Price (1993, p. 187) is that “an event A is a cause of a distinct event B just in case bringing about the occurrence of A would be an effective means by which a free agent could bring about the occurrence of B.” Menzies & Price explicate what an effective means is in terms of agent probability, that is the “probability that B would hold were one to choose to realize A” (Menzies & Price 1993, p. 190), in short ‘$\mathrm{P}_A(B)$’. Now, given agent probabilities, “A constitutes a means for achieving B just in case $\mathrm{P}_A(B)$ is greater than $\mathrm{P}_{\neg A}(B)$.” (Menzies & Price 1993, p. 190)

Price (1991) points out that a consequence of this agency theory of causation is a projectivist view of causation. For, in defining probabilities through agency, the agency theory depends on our perspective as agents. Claiming that there is a causal relation in the world amounts to projecting “onto (what we take to be) the observed world certain products of our rôle as participants.” (Price 1991, p. 173) Menzies & Price’s agency theory therefore cannot be a theory about what causation is in the world independently of us. On top of that, this agency theory is subjective in the sense that different agents might have different capabilities to manipulate the world and therefore might come to ascribe different probabilities to certain actions. This, however, is considered to be unproblematic by Menzies & Price, since, as they claim, even very different agents can “nevertheless envisage the same range of possible causal relations”. (Menzies & Price 1993, p. 200) This might be understood in the sense that different agents, even though having different actual capabilities, share the same range of possible capabilities; a claim that certainly can be doubted, but there are more serious problems.

More importantly still, Price (1991) argues that a projectivist agency theory is compatible with there being real causation in the world, and that agency causation can be "our mode of access to real causal relations." (Price 1991, p. 173) Contrary to that, Beebee (2007, p. 228) maintains:

> [It] is important to realise that, on a projectivist view, this does not involve our mistakenly *assuming* that there are mind-independent causal relations. The non-descriptive semantics of our causal talk would rule out the possibility of our even being capable of making this assumption: to *think* that there are mind-independent causal relations, in the representational sense, requires that the meaning of 'causal relations' is descriptive, which of course is what is being denied.

It is therefore questionable how far projectivism is even compatible with realism at all. At the very least, I believe that a projectivist account of causation is neither satisfactory for Canberra planners nor for naturalists, since both want to find out what causation is mind-independently. Be that as it may, I will not discuss things further here, since I think that major problems of the agency theory lie elsewhere.

An immediate objection to Menzies & Price's account is that the term 'causation' is analysed with 'to bring about', and that 'to bring about something' is just another way for saying 'to cause something'. This renders the theory circular in two places, namely when an agent brings about A and when A brings about B.[30] The answer of Menzies & Price to this charge is that 'bringing about' can be understood by ostension.

> [W]e all have direct personal experience of doing one thing and thence achieving another. [...] It is this common and commonplace experience that licenses what amounts to an ostensive definition of the notion of 'bringing about'. In other words, these cases provide direct non-linguistic acquaintance with the concept of bringing about an event; acquaintance which does not depend on prior acquisition of any causal notion. An agency theory thus escapes the threat of circularity. (Menzies & Price 1993, p. 194 f.)

Even though this reply is certainly not unproblematic and raises a number of questions, for the sake of the argument I will assume that it is sufficient to resolve the circularity.[31] I am in particular happy to grant that there is not circularity, since again I believe that there are two other problems that are more severe for this agency theory.

The first reason why Menzies & Price's agency theory fails is that it cannot account for cases of causation where there is no agent involved. The problem is that there are many processes that we usually describe as causal, but that either are not manipulated by agents even though they could, or that in practice can never be manipulated by agents. If the definiens of causation includes that it has to be an agent who brings about the cause, then why should we think that, e.g., the moon causes the tides? Menzies & Price try to solve this problem by making an analogy between causation and colour. If colour is constructed as a secondary quality, then it

[30] As Ahmed (2007) points out, there might be a third circularity involved in that 'freedom' can only be understood by reference to causal concepts.

[31] See Ahmed (2007) for a discussion of this problem.

is easy to give an answer to the question whether things have a colour even if nobody sees them. Of course they have, because all things that have a colour have intrinsic properties that have the disposition to look colourful to an observer under normal conditions. Menzies & Price transfer this line of thought to causation and argue that processes that do not involve an agent can nevertheless be regarded as causal, if they have the same intrinsic properties as causal processes involving agents:

> For we would argue that when an agent can bring about one event as a means to bringing about another, this is true in virtue of certain basic intrinsic features of the situation involved, these features being essentially non-causal though not necessarily physical in character. Accordingly, when we are presented with another situation involving a pair of events which resembles the given situation with respect to its intrinsic features, we infer that the pair of events are causally related even though they may not be manipulable. (Menzies & Price 1993, p. 197)

Now it would, of course, be interesting to know what intrinsic properties Menzies & Price refer to here, but they do not give an answer to that. This by itself makes their solution highly speculative.

However, there are more reasons to doubt that their proposal works. The agent's perspective is an essential part of Menzies & Price's theory, and they use it for example to resolve the problem that probabilistic dependence is symmetric, while causation is asymmetric. An agent can only act from the cause towards the effect and not the other way round; hence from the agent's perspective only one direction of statistical dependence makes sense. Nonetheless, on the projectivist account, this perspective is merely projected into the world and there is nothing mind-independent that the perspective would refer to. It therefore seems that there is no intrinsic property that realises the agents perspective when an agent brings about the cause, and that also would be present when there is no agent that brings about the cause, rendering the proposed way to extend the agency theory futile. A possible answer here could be that we project our perspective of agents also onto inanimate objects like the moon or stones, so that they in a sense become agents themselves. However, this possibility has been criticised by Woodward (2009), who remarks, that "we need more details about how this projection process works and why people engage in it, especially since the attributions in question, if taken literally, are so obviously mistaken – rocks are not really agents and so on."

Unrecognised by Menzies & Price, in one of their examples, they have given another possibility of how their agency theory could account for unmanipulated or unmanipulable causal processes, without invoking intrinsic properties. In arguing that the 1989 San Francisco earthquake was the effect of some unmanipulable cause, they write:

> Clearly, the agency account [...] allows us to make causal claims about unmanipulable events such as the claim that the 1989 San Francisco earthquake was caused by friction between continental plates. We can make such causal claims because we believe that there is another situation that models the circumstances surrounding the earthquake in the essential respects and does support

> a means-end relation between an appropriate pair of events. The paradigm example of such a situation would be that created by seismologists in their artificial simulations of the movement of continental plates. (Menzies & Price 1993, p. 197 f.)

This can be read as proposing the idea that unmanipulable events count as causal just in case we can model them using manipulable events. This might sound plausible, but the question when a model represents is difficult to answer. Most helpful for Menzies & Price's purposes seems to be an account of representation that does not use intrinsic properties, but instead explains representation in terms of structural similarity (cf. Bueno & French 2011). Such a theory of representation might provide the means for Menzies & Price to explain why a process is causal even though it does not involve agents, without having to rely on opaque intrinsic properties.

However, as Woodward (2009) argues, it is highly questionable whether this reasoning can succeed. First of all, Menzies & Price cannot claim that the unmanipulable process is causal, because the manipulable model represents its causal structure. This method could only work, if we already knew the causal structure of the process that is represented; only then can the relevant resemblance relation between the process and the model be defined. Since, however, the former does not involve agents, its causal structure cannot be captured by an agency theory of causation. A possible way out might be to maintain that the unmanipulable process is causal, because the model is causal and represents the process at least in some relevant non-causal way. However, according to Woodward, this reply does not work either, since it can happen that a model is representative in certain ways, without capturing the causal relations. For example, it can be the case that there are causal relations that do not appear in small-scale models, but become relevant on larger length scales of unmanipulable real world situations.[32]

A final reason for why Menzies & Price's agency account is unsuccessful is their reliance on probabilities. On their account, a cause A has to be an effective means to reach an effect B, and that is the case if and only if the probability for B is higher given A, compared to not A. I have already argued in section 3.3.3 that, even though causes might typically raise the probability of their effects, there is no general connection between causation and statistical relevance. Hence, it comes to no surprise that Menzies & Price's agency theory falls prey to some of the same objections that probabilistic theories of causation face. Even though it might be true that agency probabilities resolve some of the traditional problems of probabilistic theories of causation, like for example the problem of correlated events that have a common cause, I want to maintain that agent probabilities do not help with cases of chance lowering causation.

Consider the following situation: James is an excellent golfer, and the probability for him to hit the ball perfectly, event A, and score a hole in one, event B, is roughly 0.1. However, on this particular occasion, James makes the free decision to not hit

[32] See also Heidelberger (1992, p. 141 f.) for this point and a more extended criticism of agency causation.

the ball perfectly, thus not A is the case, which certainly lowers the chance to score a hole in one to a negligible number. (Readers who are familiar with literature on causation already know what comes next.) As it happens, a squirrel skips across the golf course and kicks the ball just in the right way so that it ends in the hole, thus B is the case. Even though James decision to not hit the ball perfectly lowered the chance for a hole in one, he scored a hole in one, and intuitively his hitting the ball not perfectly counts as a cause for this result. Yet, following Menzies & Price we would come to a different conclusion. Since not hitting the ball perfectly is not an effective means for James to score a hole in one, it cannot be the cause of him scoring a hole in one. This is clearly an unacceptable counterintuitive consequence.

Here is another example: Atom a has a probability of 0.9 to spontaneously emit radiation within one hour. Atom b only has a probability of 0.01 to spontaneously emit radiation within one hour. A scientist wants to test the effect of radiation on DNA, and using atom a would be an effective means to do so. However, the scientist freely choses to use atom b instead. As it happens, atom b emits radiation 30 minutes after the scientist started the experiment, and the DNA mutates. According to Menzies & Price, the radiation of atom b did not cause the mutation of the DNA, since the experimenter chose an ineffective means to bring it about.

Interventionism

Even though agency theories fail, one of their merits is certainly to highlight, contra Hume, that humans do not only have access to the world by observing it passively, but that they can also intervene in the processes that happen (cf. von Wright 1971, p. 82; Menzies & Price 1993, p. 191). I wish to suggest that a better way to capture this important insight is by way of an interventionist theory of causation, as developed in Woodward (2003). To avoid the problems that Menzies & Price face, Woodward relies on the notion of intervention, which, as will be shown below, is not including any reference to agents. Furthermore, he does not aim at a reductionist theory of causation. Rather, he wants to establish criteria that are sufficient and necessary to detect causal relations. Roughly, his claim is that "we see whether and how some factor or event is causally or explanatorily relevant to another when we see whether (and if so, how) changes in the former are associated with changes in the latter." (Woodward 2003, p. 14)

As Woodward understands it, causation is a relation between variables, that is, properties that can take on more than one value (cf. Woodward 2003, p. 39). The commitment to variables should, however, not be mistaken for an ontological commitment. As Woodward understands interventionism, it is not even committed to realism about causation, even though it is compatible with it. (cf. Woodward 2003, p. 121) Rather, variables allow for a convenient definition of interventionism. In particular, variables make it easy to distinguish between type and token level causation, such that a variable is on the type level, while a particular value of a variable is on the token level. Furthermore, it seems to be possible to translate propositions about variables into propositions about events (cf. Psillos 2007, p. 95), which again implies that the reference to variables is merely a convenient tool.

Woodward puts interventionism into a sufficient and a necessary condition for causation. In what follows, I will first present these conditions, before going into a bit more detail about the terms that are used in them. The two conditions are:

> **(SC)** If (i) there is a possible intervention that changes the value of X such that (ii) carrying out this intervention (and no other interventions) will change the value of Y, or the probability distribution of Y, then X causes Y.
> [...]
> **(NC)** If X causes Y then (i) there is a possible intervention that changes the value of X such that (ii) if this intervention (and no other interventions) were carried out, the value of Y (or the probability of some value of Y) would change. (Woodward 2003, p. 45)

The notion of 'cause' that is used here is that of a total cause.

> **(TC)** X is a total cause of Y if and only if there is a possible intervention on X that will change Y or the probability distribution of Y. (Woodward 2003, p. 51)

Total causes, however, are not sufficient to capture all possible causal relations, since it might happen that Y has more than one cause. In particular, it can happen that two or more causes for Y are correlated in a way that they always cancel each other out perfectly. In that case **(NC)** will break down, since an intervention on X is cancelled out by another cause and does not lead to a change of the value of Y. To rescue at least **(SC)** from this problem, it is necessary to introduce the notion of a contributing cause. Woodward does this by first defining what a direct cause is, and then using direct causes to define contributing causes:

> **(M)** A necessary and sufficient condition for X to be a (type-level) direct cause of Y with respect to a variable set $\mathbf{V}$ is that there be a possible intervention on X that will change Y or the probability distribution of Y when one holds fixed at some value all other variables Z_i in $\mathbf{V}$. A necessary and sufficient condition for X to be a (type-level) *contributing cause* of Y with respect to variable set $\mathbf{V}$ is that (i) there be a directed path from X to Y such that each link in this path is a direct causal relationship; that is, a set of variables Z_1... Z_n such that X is a direct cause of Z_1, which is in turn a direct cause of Z_2, which is a direct cause of ... Z_n, which is a direct cause of Y, and that (ii) there be some intervention on X that will change Y when all other variables in $\mathbf{V}$ that are not on this path are fixed at some value. If there is only one path P from X to Y or if the only alternative path from X to Y besides P contains no intermediate variables (i.e., is direct), then X is a contributing cause of Y as long as there is some intervention on X that will change the value of Y, for some values of the other variables in $\mathbf{V}$. (Woodward 2003, p. 59)

Equally as important for the understanding of interventionism is the definition of what an intervention is. Woodward defines an intervention, *I*, first on the type level:

> **(IV)**
> I1. I causes X.

> I2. I acts as a switch for all the other variables that cause X. That is, certain values of I are such that when I attains those values, X ceases to depend on the values of other variables that cause X and instead depends only on the value taken by I.
> I3. Any directed path from I to Y goes through X. That is, I does not directly cause Y and is not a cause of any causes of Y that are distinct from X except, of course, for those causes of Y, if any, that are built into the I-X-Y connection itself; that is, except for (a) any causes of Y that are effects of X (i.e., variables that are causally between X and Y) and (b) any causes of Y that are between I and X and have no effect on Y independently of X.
> I4. I is (statistically) independent of any variable Z that causes Y and that is on a directed path that does not go through X. (Woodward 2003, p. 98)

and then on the token level:

> **(IN)** I's assuming some value $I{=}z_i$ is an intervention on X with respect to Y if and only if I is an intervention variable for X with respect to Y and $I = z_i$ is an actual cause of the value taken by X. (Woodward 2003, p. 98)

It is now obvious, that **(SC)** and **(NC)** cannot serve as a reductive theory of causation, since interventions are defined as causes. However, this circularity is not vicious, since it only presupposes that there is a causal relation between I and X, “*but not information about the presence or absence of a causal relationship between X and Y.*” (Woodward 2003, p. 105) Thus, interventionism can still serve its purpose, namely, as a criterion to decide whether there is a causal relation between X and Y. On the flip side, this definition of interventions has no reference to agents, and therefore avoids the problems that agency theories face. **(IV)** and **(IN)** are unspecific enough so that every process in nature can potentially serve as an intervention “as long as it has the right causal characteristics.” (Woodward 2003, p. 94)

It is, however, less clear in what sense interventions have to be possible. That is, do they have to be physically, metaphysically or logically possible? First of all, it is evident that for **(SC)** and **(NC)** to be true, it is not necessary for an intervention to actually happen. Instead, it is sufficient to formulate counterfactuals about what would happen in case of an intervention. Accordingly, for Woodward (2003, p. 130 ff.) a minimal requirement on the possibility of interventions is that there are some grounds to assess the truth values of these counterfactuals.[33] This even includes that interventions might violate physical laws. For Woodward then, interventions have to be only logically possible. However, he concedes (cf. Woodward 2003, p. 112) that there can be cases in which it is not meaningful to change a property or in which we have no possibility to find out what would happen under interventions. In these cases the interventionist theory will not be applicable. I currently do not see any way how to formulate a general rule for when the notion of change of a value is meaningful and the truth value of relevant counterfactuals can be found. Rather, it seems that this has to be decided on a case to case basis.

[33] See Psillos (2007) for a critical assessment of the evaluation of counterfactuals within interventionism.

This is the starting point for a common criticism on interventionism. The prime example here is that of the whole universe at time t_0 causing the whole universe to be in a certain state at the later time t_1. Let us call this a case of immanent causation:

(IC) If an entity exists at a time t_0 and this entity also exists at a later time t_1, then this entity at t_0 is a sufficient cause for the effect that is its existence at t_1.

The problem is that by definition interventions have to be exogenous, that is, they have to be outside of the X-Y system. If now in the example X and Y are states of the entire universe, then it seems impossible that X can be intervened on from the outside. Even though Woodward calls this an "unresolved issue", the solution that he suggests is to limit the scope of interventionism to a smaller scale: "we should take seriously the possibility that causal reasoning and understanding apply most naturally to small world systems of medium-sized physical objects of the sort studied in the various special sciences". (Woodward 2009, p. 258) Woodward motivates this move by pointing out that it is a widespread belief that there is no causation in physics. Thus, when the state of the universe is taken to be its fundamental physical state, we should not expect that any theory of causation applies to it. I believe, however, that for the following reason Woodward's reply is overtly ad hoc. The problem does not hinge on that the state of the universe is described by fundamental physics. One might as well describe the universe as composed out of everyday objects, and still find **(IC)** true of the whole universe. Consequently, the strategy to limit the applicability of interventionism to small worlds seems entirely motivated to avoid the present problem and is therefore not very convincing.

I want to suggest a different solution to the problem of the unmanipulable universe, which is that objects should never be understood as the cause for their existence at later times in the absence of any interaction; in short, there is no immanent causation. According to Salmon (1980a) there are two basic causal concepts, namely production and propagation, and although they are "intimately related to one another, we should, I believe, resist any temptation to try to reduce one to the other." (Salmon 1980a, p. 50) The difference between them is that causes produce their effects via a causal process. Or, in other words, "causal processes, in many, instances, constitute the causal connections between cause and effect. A causal process is an individual entity, and such entities transmit causal influence." (Salmon 1980a, p. 66) This does not mean that every causal process has to form a relation between cause and effect, but only that it has the potential to do so. What matters here most is that the distinction between production and propagation is essential and neither should be eliminated. If causation is understood as a relation, then there need to be distinct relata and a relation between them (cf. section 3.3.1 above). If, however, as in **(IC)** an entity that persist through time is a series of causes and effects, then the relation is reduced to a series of infinitely many relata, one at every instant of time. As a consequence, it becomes opaque what constitutes the relation between cause and effect, and indeed, it seems that accepting **(IC)** amounts to a considerable change of the concept 'causation' – a harmful change since it is not clear what the alternative concept is.[34] Thus, I content that there is no immanent causation and causal processes should not be understood in terms of **(IC)**. To wrap up this argument, since interventionism is defined through causes and effects and the persisting universe should not be understood as a series of causes and effects, the persisting universe is not an entity to which interventionism should apply.

A possibly reply to the dismissal of **(IC)** might be that it still counts as a causal relation according to Woodward's interventionism. For the sake of the argument, I will forget the above mentioned conceptual problems for a second. If now X=S(t_0) is the state of an object (smaller than the universe) at time t_0 and Y=S(t_1) is the state of that object at time t_1, it clearly is possible to manipulate S(t_1) by manipulating S(t_0) and thus they are causally related, so one might argue. Hence, according to interventionism an object can be the cause of itself at a later time. Nevertheless, for the following reason, I think that it is at least highly dubitable that interventionism applies to **(IC)**. The first question to ask is what kind of process would count as an intervention on S(t_0) in this case. It is crucial to recall condition **(IV)**.I2, that is, an intervention on X has to break all other causal connection that previously determined the value of X. Now, if S(t_0) is the complete state of an object, then according to **(IC)** one of the causes of S(t_0) is S($t_{0\text{-}1}$), that is, the state of the object before t_0. How can an intervention break the causal connection between S(t_0) and S($t_{0\text{-}1}$)? There are only two ways; either the intervention brings the object into existence at t_0, or the intervention puts an end to the existence of the object at t_0. In both cases the intervention has an obvious consequence for S(t_1). Interventions of another kind, say, to change the velocity of the object will not suffice, since S(t_0) with a changed velocity will still be immanently causally depend on S($t_{0\text{-}1}$). If this is correct, then the only way to manipulate S(t_1) is to manipulate the existence of S(t_0). However, it is at least questionable whether existence is a suitable variable for interventions. It seems that a minimal requirement for the latter would be to accept existence as a real predicate that objects have. For some philosophers this might be an acceptable view, but it is certainly not an uncontroversial one that is easily embraced to rescue **(IC)**.[35]

Menzies (2004, 2009) has argued for a way to amend interventionism by strengthening the intuition that "a cause is an intervention that makes a difference *to the*

[34] As Salmon (1980a, p. 67) remarks, it nevertheless might be possible to change the concept of causation in that way: "I see no reason for supposing that this program could not be carried through, but I would be inclined to ask why we should bother to do so. One important source of difficulty for Hume, if I understand him, is that he tried to account for causal connections between non-contiguous events by interpolating intervening events. This approach seemed only to raise precisely the same questions about causal connections between events, for one had to ask how the causal influence is transmitted from one intervening event to another along the chain."

[35] Eagle (2007, p. 158) points to another argument against the example of the whole universe as a cause or effect, similar to Woodward's reply. As he stresses, it seems that the folks understanding of 'causation' was never meant to apply to things like the whole universe, but rather orientates itself on mid-sized objects of everyday life. Given that the folk notion is what is analysed, it thus might be unfair to utilise the whole universe as a counterexample.

normal course of events." (Menzies 2009, p. 358) If again the cause is represented by the variable X and the effect by the variable Y, then the normal course of events is represented by their values x' and y' respectively, or accordingly by the pair $<X=x', Y=y'>$. Adding this to Woodward's interventionism, Menzies comes up with the following variation of it:

> $X=x$ causes $Y=y$ relative to the context $<X=x', Y=y'>$ if and only if (i) the actual values of X and Y are x and y respectively; and (ii) if an intervention were to change the value of X from x' to x then the value of Y would change from y' to y. (Menzies 2009, p. 360)

As Menzies points out, in contrast to **(SC)**, to determine the truth value of counterfactuals in his interventionist framework, we need to know what the normal course of events is, and this according to Menzies very often depends on normative or otherwise mind-dependent considerations. As a consequence, Menzies (2004, p. 139) urges "to give up the metaphysical assumption that causation is an absolute relation, specifiable independently of context." Menzies' argument for this alteration is that it is supposed to resolve several well-known puzzles about causation, e.g., how causation by absence should be understood. This is not the place to assess the validity of Menzies' claim. Suffice it to say that I believe that, at least in he context of physics, the problems that Menzies refers to can be solved on a different way. But more important here is that Menzies does not introduce his theory because of a defect of interventionism. **(SC)** and Menzies criterion agree on which variable X will cause which variable Y. Thus, even is Menzies should be correct, it is not the case that **(SC)** identifies false causal relations.

It is, however, noteworthy that Menzies and Woodward estimate the scope of their theories quite differently. While Menzies derives consequences for our understanding of what causation is in the world, Woodward often stresses that his interventionism is merely "a nontrivial *constraint* on what it is for a relationship to be causal without providing a reductive analysis of causality." (Woodward 2003, p. 106) In this matter, I want to side with Woodward. Thus, within the methodological framework the present work follows, interventionism gives no information on what the reduction basis *is*. Interventionism merely helps to identify causal relations in the world, and imposes the constraint that whatever the concept 'causation' can be reduced to, it has to fulfil its conditions.[36]

3.3.5 Asymmetry and directionality

Hausman (1998, p. 1) lists eleven different ways in which causation is an asymmetric relation. Not all of these ways are independent from one another, and some can possibly be reduced to others. In particular, I want to suggest that there are only two basic ways in which causation is asymmetric: temporal asymmetry and causal asymmetry. It seems that the intuition that causation is asymmetric in these two

[36] It should be mentioned that in the light of this role of interventionism, its circularity is particularly harmless, since interventionism is not meant to reduce causation to non-causal concepts.

ways is shared by most and is thus one of the core platitudes on causation that can hardly be repudiated.[37]

The first platitude can be put in the following phrase:

(TD) Temporal directionality: Causes come before their effects, but not the converse.

To understand this statement better, in particular to understand why it is about temporal *directionality* and not temporal *asymmetry*, it is worth to follow Reichenbach (1956, ch. 3), in stressing the distinction between order and direction. If we imagine a series of events as occurring on a line, then they stand in an asymmetrical relation to one another, that is, if event A is on the left of event B, then B is not on the left of A. This, however, has to be distinguished from the direction that series of events may have. As Reichenbach points out, order and asymmetry are not enough to establish direction. The reason is that the series of events has no internal properties that would allow distinguishing between left and right. Or in other words, the series of events in which A is on the left of B and the reverse in which B is on the left of A are structurally equivalent. We can, however, give the series a direction if we, say, draw a line on paper and define left and right with respect to that line. For example, we can give event A the name 'discovery of America' and event B 'French revolution'. If we further define that 'discovery of America' is on the left of 'French revolution', then the series has a direction, since the case in which A is on the left of B is structurally different from the case in which B is on the left of A. Reichenbach gives the instructive example of the negative and positive real numbers. Positive numbers are larger than negative numbers, which can be defined by the fact that the square of a negative number is a positive number while the square of a positive number is also a positive number. Defined that way, the relation 'larger than' cannot simply be reversed like the relation 'to the right of', and therefore is not only asymmetric, but also directed. The reason is that negative and positive numbers are not structurally equivalent. The thrust of the asymmetry-directionality distinction is that directionality is much harder to come by; any ordered series of events is sufficient for asymmetry, but not for directionality.[38]

What does **(TD)** imply about asymmetry and directionality? The first observation is that the relation referred to, 'to come before', is only a temporal relation. Thus, **(TD)** is only a statement about the temporal relation between cause and effect, but not about the causal relation between them. Furthermore, 'before' is certainly an asymmetric relation in that if A comes before B, then B does not come before A. For this reason, a statement like **(TD)** is often referred to as the temporal *asymmetry* of causation, e.g. by Price (1996b, p. 133) and Farr & Reutlinger (2013). This name, however, is incomplete at best, and misleading at worst, for **(TD)** does not only imply asymmetry, but also directionality. If causes always come before their effects, as **(TD)** implies, then the temporal relation they stand in cannot be reversed. There

[37] Though, there are exceptions such as Ney (2009) who is arguing for a symmetric causation. I will discuss Ney's theory in chapter 4.

[38] For a precise definition of order and direction with respect to time see Earman (1974, p. 18).

must be a structural difference between the case where a cause comes before its effect and the converse. A series of causes and effects therefore has a temporal direction. Or in short, for **(TD)** to be true, time must have a direction, even though this must not necessarily be a direction of time itself, that is, an anisotropy of time.[39]

The situation is similar with the second platitude of this section:

(CD) Causal directionality: If A causes B, then B does not cause A.

First and foremost **(CD)** establishes that causation is an asymmetrical relation. As before, this lead many philosophers to *only* conclude that causation is an asymmetrical relation – an incomplete conclusion, as I want to maintain, for it is certainly part of commonsense that causation also has a direction (cf. Reichenbach 1956, p. 27). However, unlike **(TD)** where the directionality is explicated with respect to time, **(CD)** does not define causal directionality via an arguably non-causal concept. One can therefore object that **(CD)**, as it stands is unilluminating, maybe even tautological, and in need for further analysis. This analysis might be provided by statements such as "The cause compels the effect in some sense in which the effect does not compel the cause." (Russell 1913, p. 10), "When we say that lightning starts a forest fire we mean that the electrical discharge produces ignition." (Salmon 1980a, p. 49) or "Causes bring about effects and not *vice versa*." (Norton 2007, p. 36). Nevertheless, none of these statements really seem to give an analysis in which sense causation has a direction, since 'compel', 'manipulate', 'produce' and 'bring about' in this context are more or less only different words for 'cause'. Even if this is not made clear or not intentionally implied by most authors cited here, I want to maintain that intuitively **(CD)** not only is a statement about asymmetry, but also about the directionality of causation. If A causes B, then this is more then a mere convention, and a series of causally related events cannot be reversed, that is, it is unidirectional.[40]

It is instructive to point out the difference between causal asymmetry and the asymmetry of implication. As Simon & Rescher (1966, p. 323) explain, an asymmetric implicational proposition like 'If one has food poisoning, one also has stomach ache' can be contraposed into 'If one not has stomach ache, one also not has food poisoning'. Contrary to that, a causal statement like 'Food poisoning causes stomach ache' cannot be contraposed into 'A pain free stomach causes not having food poisoning'. Simon & Rescher (1966, p. 324) conclude that the "asymmetry of the causal relation is unrelated to the asymmetry of any mode of implication that contraposes." Hence, if a suitable reduction basis for causation is sought, it is not sufficient that this basis comprises implicational relations between events. For example, that the force, acting on an object, and the object's mass imply its acceleration is not sufficient to infer that they cause the objects acceleration.

It seems to be a commonsensical belief that the causal direction is always aligned with the temporal direction. The latter is also implied by a conjunction of **(TD)**

[39] I will present a more detailed discussion of this point in chapter 4.
[40] Again, more on this will be said in chapter 4.

and **(CD)**; if A causes B, then A comes before B in time. This in turn prompts the question whether the causal direction can be reduced onto the temporal direction. Price (1996b, p. 137) ascribes such a reductionism to David Hume. Accordingly, the causal relation is symmetrical between cause and effect and if A and B are causally related, then by definition whatever comes first in time is the cause and the other event is the effect. There are, however, reasons to reject such a reductionism (cf. Lewis 1973, p. 170; Earman 1976, p. 10 f.; Price 1996b, p. 137 f.). First, that 'cause' is often replaced by terms like 'produce' or 'bring about' suggests that intuitively causation itself is a directed relation and that there is more to the distinction between cause and effect than merely a stipulation with respect to the temporal order of events.[41] Second, this reductionism would make it a priori true that there cannot be simultaneous or backwards causation. Whether there is one of the latter or both, however, should be an empirical rather than an analytic question. And finally, as some philosophers, e.g. Reichenbach (1956), have argued, the reduction should go in the other direction, that is, the direction of time should be reduced to the direction of causation. Accordingly, there has to be at least a further argument to the effect that one direction of reduction rather than the other is correct. This is not the place to go further into the discussion about reduction, so for now I only wish to suggest that intuitively causation is directed in two ways, temporally and causally.

As a final remark, I would like to point out that **(TD)** itself excludes the possibilities of simultaneous and backwards causation which, again, is rejected by many as an unjustified ruling of philosophy over science. I do indeed believe that **(TD)** is the correct expression of the folk's intuition on causation. Nevertheless, from the previous sections it should be clear that this platitude does not have to come out as true in all cases. If empirical reasons show up that suggest that there is simultaneous or backwards causation, as might happen especially in quantum entanglement (cf. Corry 2015), then I would expect that **(TD)** can be adjusted to that without difficulty, most likely by adding a ceteris paribus clause that lists cases in which effects temporally precede their causes. At the moment, I do not see any compelling reason to believe that simultaneous or backwards causation exist, but a reason might very well be found in the future.[42]

3.3.6 The law of causality

An often expressed verdict on causation is what Norton (2007, p. 36) calls the 'constancy of causation': '*'the same causes always bring about the same effects*." Russell (1913, p. 6 f.) criticises this 'law of causality', as he calls it, in two ways.[43] First of all, if events are sufficiently fine grained, especially when they are described

[41] This example shows also that Farr & Reutlinger (2013, p. 217) are mistaken in claiming that "the failure of causal asymmetry entails the failure of causal time asymmetry (since the latter requires the discernibility of cause and effect)". Contrary to their opinion, causation can be a symmetric relation while the distinction between cause and effect is given by their temporal order. (cf. Reichenbach 1956, p. 29)

[42] See Earman (1976) for a discussion of backwards causation.

[43] More recently, Russell's criticism was endorsed by Hitchcock (2007, sec. 3.3).

by physics, then it is extremely unlikely that two events happen that are exactly the same. This by itself does not falsify the law, but it turns it into an empty law without instantiation. On the other hand, this is of course already sufficient to support Russell's claim that causation is not a part of mature science. Russell's second objection is that it is in principle always possible that an intervention occurs, which prevents the effect from happening. Thus, even if, almost per impossibile, in the course of the universe two events occur that are exactly the same it does not follow that in both cases the same effect, or any effect has to occur.

Nevertheless, Russell concedes that the law of causality seems to be true in a weaker form: "In spite of these difficulties, it must, of course, be admitted that many fairly dependable regularities of sequence occur in daily life." (Russell 1913, p. 8) Armstrong (2004, p. 454 f.) goes in the same direction, when he counts 'same cause, same effect' as a basic platitude about causation and explains: "I mean by this that it is regularly true, even allowing that there are many apparent exceptions in ordinary experience, that from the same sort of cause the same sort of event, or at least roughly the same probability distribution of events, will follow."[44] A more detailed explication of this weak law of causality can be found in Woodward (2003) under the notion of 'invariance':

> The guiding idea is that invariance is the key feature a relationship must possess if it is to count as causal or explanatory. Intuitively, an invariant relationship remains stable or unchanged as various other changes occur. Invariance, as I understand it, does not require exact or literal truth; I count a generalization as invariant or stable across certain changes if it holds up to some appropriate level of approximation across those changes. (Woodward 2003, p. 239)

Accordingly, if there is a causal connection between X and Y, then this should continue to hold, i.e., be invariant, even when X is changed. Also, it is irrelevant whether the change occurs due to an intervention in the strict sense. Furthermore, it is admissible that there are some interventions under which the causal connection breaks down (cf. Woodward 2003, p. 241 ff.). For example, when there is a causal connection between the temperature a thermometer measures and the outside temperature, then we can expect this connection to hold for a variety of different outside temperatures. At the same time, there might be temperatures where the connection breaks, e.g., because they are high enough to exceed the scale of the thermometer.

Russell's notion of dependable correlations can be understood as realised by invariant connections, since invariance makes a connection reliable in different circumstances. Analogously, Armstrong's intuition that the same sort of effect follows from the same sort of cause is captured by invariance, if we understand an event sort roughly as, speaking with Woodward, the set of values of a variable for which the causal connection holds.

[44] See also Ross & Spurrett (2007, p. 55): "We agree with Russell — and with most current philosophy of science — that scientists do not generally seek 'invariable uniformities of sequence'. But it does not follow from this that they seek nothing worth calling causes."

It is worth noting that invariance, as Woodward understands it, is essentially equivalent to what Redhead (1987, p. 103) calls ‘robustness’: “The requirement of robustness as a necessary condition for a causal relation means that suitably small disturbances of either relata do not affect the causal relation.” Thus, while invariance points out that causal connection should continue to hold under interventions, robustness means that, given the interventions are sufficiently small, the causal connection will continue to hold and on top of that any change of the causal relata is probably negligible.

All explications of the weak law of causality presented here are quite informal and leave plenty of space for unsharp boundaries between causal connections and mere correlations. There might be extreme cases in which there is a causal connection that breaks down under almost every intervention, and conversely there might be a correlation that is extremely stable under interventions. In these cases we will have to find other criteria to decide whether a correlation is causal or not. Nevertheless, there certainly are very clear instantiations of the weak law of causality. The take away message of this section’s platitude may therefore be formulated as that it can be expected of a causal connection to be invariant under a range of interventions. Even though invariance is neither a sufficient nor a necessary condition for causation.

3.4 Conclusion

From the apparent insufficiencies of previous transference theories of causation, as shown in chapter 2, the necessity for a systematic methodological approach to causation grew. In section 3.2, this method was found at the intersection of the Canberra plan and naturalism. It has been argued that the method of choice, to answer the questions whether there is causation in physics and what it is, is to first perform a conceptual analysis in finding platitudes that are commonly associated with causation and second to investigate whether a conjunction of those platitudes is realised in physics. The most part of this chapter has been devoted to the first step, that is, finding and explicating the platitudes. The second step of the methodology will be pursued in the remainder of this work.

The main results of this chapter are the following six platitudes:

(P1) *Relation and Relata:* Causation is a relation between distinct events.

(P2) *Absences:* Absences do not occur as causal relata in physics.

(P3) *Probability:* Causes typically raise the probability for their effects.

(P4) *Interventionism:* A theory of physical causation should agree with interventionism.

(P5) *Direction:* Causation is temporally and causally directed.

(P6) *The law of causation:* Causal connections typically continue to hold under interventions.

It should be noted that (P2) and (P3) are largely negative, that is, neither do absences have to be incorporated in a theory of physical causation, nor is there a strong connection between causation and statistical relevance. Also, I wish to point out that (P4), (P5) and (P6) are somehow odd platitudes, compared to the others, in that they do not provide an analysis of 'causation' into more basic terms, but provide constraints that a theory of causation has to fulfil. This, however, does not prohibit (P4) (P5) and (P6) from being part of the folk theory of causation, since they provide valuable help in singling out possible realisers of the folk theory and exclude others. Finally, from the list of platitudes the following condition can be formulated, that a physical thing should realise to be called causal: *Causation is a relation between distinct events that can be exploited for manipulation, it has a direction and is stable under interventions.*

I cannot claim that this list of platitudes is complete and satisfies all usages of the term causation, and I do not believe that this is even a realistic goal. To say it with Reichenbach (1956, p. 24):

> An explication can never be proved to be strictly correct, for the very reason that the explicandum is vague and we can never tell whether the explicans matches all its features. We can merely require that an explication be adequate, that is, that the explicans correspond, at least qualitatively, to the usage of the term in conversational language, and that if the explicans is put into the place of the explicandum, most sentences of conversational language do not change their truth values.

If the platitudes that were explicated here together form a concept that generally agrees with commonsensical usage and has sufficient meaning to be informative, that is, to single out some kind of process in the world and exclude others, then the aim of this chapter is reached. It will be the burden of subsequent chapters, nonetheless, to find out whether the above platitudes are realised in physics.

4 (A)symmetry in causation and physics

4.1 Introduction

Even though the analysis of the folk's intuition about causation has come to a sufficiently clear result with the finding of several platitudes, it is not yet demonstrated whether these platitudes are realised by physics. Unsurprisingly, every one of the six platitudes has been challenged as being incompatible with modern physics. But probably the most historically influential debate centres around the topic of symmetry and asymmetry.[1] I will therefore begin the argument for a causal interpretation of physics by analysing whether and how directed causation is congruent with the apparent symmetries of physics. For the most part, with an exception in section 4.3.1, the discussion will not directly be based on QFT, which is why I choose to put it before the detailed exposition of the formalism of QFT.

One of the first to observe that the old concept of causation hardly accords with modern science was Ernst Mach. According to Mach:

> [T]he old-fashioned idea of causality is a little clumsy: A dose of cause is followed by a dose of effect. This represents a kind of primitive, pharmaceutical *Weltanschauung*, like in the doctrine of the four elements. [...] The connections in nature are rarely ever so simple that one can identify *one* cause and *one* effect. Consequently, I have tried for a long time to replace the concept of cause by the mathematical concept of the functional relation: dependency of the appearances from each other. (Mach 1886, p. 74, cited after Heidelberger 2010, p. 171)[2]

However, it was famously Bertrand Russell (1913) who formed the concerns of Mach into arguments which have dominated the discussion ever since. Following to Russell, there are two symmetries of physics that contradict causation. First of all, physics is

[1]It is commonplace in the recent literature to speak about the asymmetry of causation, rather than the directionality. It will sometimes be unavoidable to follow this usage, however, as I have argued in section 3.3.5, I still want to maintain that causation is directed. Even though asymmetry implies that when a causes b, i.e. Cab, it cannot be true that Cba, but asymmetry does not preclude us from reversing the relation C such that now Cba and not Cab. The directionality of causation is meant to imply that the reversal of causation is not again a causal relation, or in other words, if a causes b, then this relation cannot be reversed.

[2]It is historically interesting that, as Heidelberger (2010) points out, in 1848 Emil Du Bois-Reymond still saw functions as describing causal connections, and that even later Gustav Theodor Fechner believed both to be compatible at least. See also Scheibe (2007, ch. VII) for a historical overview.

time reversal invariant, that is, it is temporally symmetric, which seems to stand in opposition to the platitude that the cause comes before the effect. Second, physics is *nomically bidirectional*, such that if an event A determines an event B, then B also determines A, which seems to contradict the platitude that if A causes B, then B does not cause A. As a consequence of these two and other arguments, like Mach, Russell concludes that modern physics has replaced the old fashioned concept of causation with that of a functional relation.

According to Farr & Reutlinger (2013, p. 218), in the recent discussion the insight of Russell has been put into what they call the directionality argument:

1. *If* the dynamical laws of fundamental physics are causal laws, *then* they are time-asymmetric [...].
2. The fundamental physical theories/laws are time-symmetric.
3. Therefore, the fundamental physical theories/laws are not causal.

As Farr & Reutlinger point out, most authors who support this argument do not make sufficiently clear what 'time-symmetric' is supposed to mean, that is, whether it refers to the time reversal invariance or the nomic bidirectionality of physics. I agree with Farr & Reutlinger that in order to answer the challenge of the directionality argument it is crucial to keep these two meanings distinct. Accordingly, in this chapter I will first discuss the directionality argument based on the time reversal invariance of physics. In agreement with Farr & Reutlinger, the result will be that the argument in this reading is unsuccessful. I will then continue to show that also the directionality argument based on nomic bidirectionality fails.

The failure of the directionality argument from time reversal symmetry does not imply that physics in turn supports the notion of a temporal direction of causation. It could still be the case that physics is temporally symmetric, but that this symmetry does not contradict causation. In this way, there would be no possibility to ground the temporal directionality of causation in physics, even though both would be compatible. Consequently, I will start the second half of this chapter arguing that even time reversal invariant laws can give rise to temporally asymmetric processes. In particular, I will argue that symmetric laws are compatible with asymmetric boundary conditions, and, following Castagnino & Lombardi (2009), that a direction of time can be defined from the local flux of energy. Even though this directionality will not make it an objective fact which direction of time is future and which is past, I will suggest that an objective distinction between future and past can be established and that this is sufficient to ground the temporal direction of causation.

I will then continue to argue that the other formulation of the directionality argument, based on nomic bidirectionality, can be rebutted, if the intimate connection between causation and intervention is taken into account, and determinism is understood in a model relative sense. That interventions comprise some sort of asymmetry has been invoked as an argument several times already, most recently in Frisch (2010). The sort of asymmetry employed here is temporal asymmetry in the form that interventions come before their effects in time. Using the temporal asymmetry of interventions as an argument has, however, already been criticised by Reichenbach (1956, ch. 6). I will first explain in more detail, why I agree with Reichenbach on this point. Subsequently, I will argue that there is a different sense in which interventions involve a certain kind of asymmetry, namely an asymmetry of nomic dependence. Building up on Farr & Reutlinger's recent result that the main thrust of the directionality argument rests on nomic bidirectionality, contrary to what is commonly assumed (see e.g. Hoefer 2004), I will show how interventions break nomic bidirectionality and causation can be reconciled with physics.[3]

4.2 The *in*compatibility of causation and physics

4.2.1 The temporal directionality argument

The first strike against causation in physics may be found in Russell (1913, p. 15), where he wrote:

> The law makes no difference between past and future: 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 other variable is a function of them.

Farr & Reutlinger (2013) interpret this passage in the sense that Russell is making a claim on the incompatibility between physics and the time asymmetry of causation. It is not overtly clear what feature of physics exactly Russell refers to that lets him believe that there is no difference between past and future, but taking it to be the time reversal symmetry of physics seems possible at least. Sticking to Russell's terminology, one could say that in a time symmetric theory the future can be a function of the past just as well as the past a function of the future, and thus that there is no difference between them.

More recently, this argument against a causal interpretation of physics has been stated, though not endorsed, most clearly by Field (2003) and Frisch (2012). For example, Field's version of the argument can be summarised with the following claims:

> (1) "The relation between cause and effect is supposed to have an important temporal asymmetry: causes normally or always precede their effects."
>
> (2) "But at the level of fundamental physical law, it is hard to see any grounds for the evident directionality of causation. The point is sometimes put a bit contentiously, by claiming that (perhaps with a few minor exceptions) the fundamental physical laws are completely time symmetric."

[3] A different way to address this problem is the neo-Russellian strategy, according to which there are higher-level causal relations in non-fundamental sciences that are dependent on non-causal physical relations (see the short survey in Reutlinger 2014). I cannot comment here on whether the neo-Russellian strategy is effective. In any case, this chapter presents an alternative.

(3) "If so, then if one is inclined to found causation on fundamental physical law, it isn't evident just how directionality gets in." (Field 2003, p. 436)[4]

According to Frisch (2012, p. 314), "the most telling contrast is supposed to be that between the asymmetry of the causal relation and the putative time-reversal invariance of the dynamical laws of our mature physical theories." And Price & Weslake (2009, p. 416) ask the question: "Fundamental physics seems to be time-symmetric, in the sense that if it permits a process to occur in one temporal direction, it also allows it to occur in the opposite temporal direction. How could time-symmetric physics yield something as time-asymmetric as the cause–effect distinction?"

As Farr & Reutlinger point out, however, the case is not as watertight as it might seem at first glance. Recalling platitude **(TD)** from section 3.3.5, the temporal directionality of causation amounts to that the cause comes before the effect, but not the converse. Furthermore, in the quotations above the time reversal symmetry of physics is understood in the sense that every physical process can happen in both temporal directions (a more precise characterisation will be given in the next section). This opens up the following loophole for a causal interpretation of physics. In positive time it might, for example, be true that lightning usually comes before bush fires and also cause bush fires. In negative time, however, the reverse is true and bush fires usually come before lightning. From here it does not follow that in negative time still lightning causes bush fires and therefore contradict **(TD)**, for it is at least possible to have the temporal directionality of causation fixed to the direction of time such that when the temporal direction reverses also the causal direction does. Hence, one can argue that while in positive time lightning causes bush fires, in negative time bush fires cause lightning. It might seem odd to claim that causal connections can hold in the opposite direction of what we are used to from everyday life, but at least it does prevent a contradiction of **(TD)** in the light of time reversal invariant physics.[5]

Notwithstanding, a thought experiment similar to that presented in the last paragraph to underpin, and not to refute the directionality argument, has been presented by Norton (2009, p. 481 f.). Norton asks the reader to imagine a world that consists only of two processes. Process A, which can be a scattering event with an initial incoming wave and a scattered wave, and process B that is isomorphic to A, but also temporally reversed. According to the principle of the temporal directionality of causation, for process A the initial wave comes before the scattered wave and causes it. What about process B? According to Norton (2009, p. 482), "using the time order natural to process A, we would have to say that the principle of causality requires the present states of process B to depend upon its future states."

[4]Field (2003, p. 436) is quite cautious and ads that "it is not obvious that the claim that the basic laws of physics are time-symmetric is correct; indeed, the notion of the time symmetry of a law itself is not as clear as it sounds." Indeed, Field favours the argument from nomic bidirectionality of physics against causation.

[5]I have more to say in section 4.3.1 on how the temporal and the causal direction can be tied together, without defining one by the other.

Norton concludes that this is a reductio ad absurdum for the temporal directionality of causation, since if it applies to process A that causes come before the effects it does not apply for process B and vice versa. However, I do not think that Norton's argument is conclusive, for prima facie there are two possible ways out. First, one might follow Norton and evaluate both processes using the same temporal direction. In that case, Norton does not explain why one cannot conclude that in process B it is the scattered wave that causes the initial wave. On the other hand, Norton does not give any reasons why, when assessing the causal relations of process B, one should 'use the time order natural to process A'. Thus, Norton gives no reason that would preclude the possibility that with the temporal direction of process B also the causal direction is reversed, such that for both processes it is true that the initial wave comes before the scattered wave and causes it, even though the processes have opposite temporal directions.[6]

A stance on causation that would make this move at least implausible, if not impossible, would be hyperrealism about causation. If there is some property that makes a cause a cause and an effect an effect over and above all physical properties, then why should the causal direction be attached to the temporal direction? If in positive time lightning has something that bush fires do not have and that makes them causes, then why and how should this something go over to bush fires in negative time? It would follow that independently of the direction of time lightning causes bush fires, and that they, say, in positive time come before bush fires, but in negative time come after bush fires. However, for several reasons hyperrealism does not seem like a desirable position to take (cf. Price & Weslake 2009). In particular, in this context one might wonder why a hyperrealist should worry about the directionality argument at all. If causation is that much removed from physics, then any concern about reconciling physics and causation seems obsolete.

As a consequence of the directionality argument's failure in the above interpretation of the time symmetry of physics, the question automatically comes up whether there is a different interpretation in which the argument succeeds. Even though in the above thought experiments time reversibility might have led to two different causal processes, there was no contradiction, since nothing demanded that both processes needed to have the same causal order. As a consequence, Farr & Reutlinger (2013) note that for the directionality argument to be valid the time reversal invariance of physics must be interpreted along the following lines:

(¬TD) *No temporal directionality*: If in one time direction it is true that event A causes event B and A comes before B, then, by the time reversal invariance of physics, it is also true that the *same* event B comes before the *same* event A.

This allows for a valid reformulation of the directionality argument:

1. If a process is causal, then, by **(TD)**, it has a unique temporal direction.
2. If a process is a physical process, then, by **(¬TD)**, it has no unique temporal direction.

[6] Also see Frisch (2009b, p. 492) for a quite different reply to Norton's argument.

3. If a process is a physical process, then it is not a causal process.

In the next section I will discuss more carefully what time reversal invariance is, which interpretation of it supports **(¬TD)** and why I do not think that it is feasible. But before, I wish to briefly present a discussion between Norton (2007, 2009) and Frisch (2009b,a), which could give reasons to believe that, independently of time reversal invariance, physics is temporally asymmetric.

Frisch (2009b, p. 461) argues for "the claim that asymmetric causal notions play a role in theorizing in physics." His argument rests on a case study on how dispersion relations are derived in standard physical textbooks. For Frisch, the particular step in the derivation is the condition that if an interaction happens at a time t_0, then it does not have any effect on the waves that participate in that interaction at times $t < t_0$ (cf. Frisch 2009b, p. 463). Frisch observes that physicists often refer to this condition as the 'causality condition' and justify it as being in agreement with the intuition that the cause comes before the effect. Furthermore, Frisch speculates that there is an underlying reason for why physicists think of this as a causality condition that draws on the connection between causation and the fact that we can only intervene on the future.

Frisch himself refrains from inferring any further conclusions in the direction that the physical processes involved in dispersion are causal themselves:

> But I do not here want to take sides in the debate as to what metaphysical conclusions we should draw from the fact that asymmetric causal relations play a substantive role in physical theorizing. Theorizing in physics involves appeals to causal constraints, just as it involves positing quarks or electrons, but what metaphysical commitments follow from this is a question I do not want to address here. (Frisch 2009b, p. 461)

Nevertheless, one could take Frisch's observations as the starting point for an argument that at least in some areas of physics temporally asymmetrical conditions are indispensable and that the temporal directionality of causation can be based on them. However, as Norton (2009) points out, the asymmetry of dispersion relations does not solve the problem that the underlying fundamental theory, namely, electrodynamics is still temporally symmetric and that dispersion relations cannot be rigorously derived from it.

> Physicists developing dispersion theory arrive at a general result through a complicated mix of intuitions about how electrodynamical systems behave, the experimental evidence, and more precise computations on artificial examples. They have sought to legitimize that result by appeals to 'causality', apparently believing that this calls up a greater body of theory capable of grounding their inference. (Norton 2009, p. 477)

Norton argues, correctly I believe, that the main motivation for physicists to impose an asymmetrical condition is to arrive at a model that is empirically adequate, and not to agree with intuitions about causation; a point that Frisch (2009b,a, p. 490, 471) seems to agree on. On top of that, Norton (2009, p. 478) stresses that what

Frisch calls the 'causality condition' is merely the choice of initial conditions. By choosing certain initial conditions, we choose one model out of a set of models that is compatible with the equations of motion. Thus to get the temporal asymmetry it is not necessary to apply a causally condition that lies outside of the theory.[7] It follows, that dispersion relations do not seem to give any additional reason to assume that fundamental physical processes are causal. Hence, in his reply to Norton, Frisch (2009a, p. 491) concedes that "Norton is correct in claiming that causal notions are not fundamental in the sense that they play no role even in our classical micro-theories of the 'inner constitution of bodies'."

4.2.2 What time reversal symmetry amounts to, or why the argument fails

For the directionality argument to go through on the basis of the time reversal invariance of physics **(¬TD)** must be true. This section will discuss whether this is the case under any feasible interpretation of time reversal invariance. To begin with, it will be necessary to spell out some technical foundations for possible interpretations.[8] I wish to stress that it is not the purpose of this section to argue that time has a direction in the sense that it is an objective fact what the future and what the past is. As the discussion will show, the latter need not be true in order for **(¬TD)** to be false.

The first step is to define in what sense there can be a temporal orientation in physics:

> A relativistic space-time $< M, g, \nabla >$ [with M being a differentiable manifold, g a Lorentzian metric and ∇ a unique symmetric linear connection] is said to be temporally orientable if there exists a continuous nonvanishing vector field on M which is timelike with respect to g. (Earman 1974, p. 17)

As Earman points out, it is not always the case that such a vector field exists and thus not every space-time is temporally orientable. We can now go on and define what it means that an orientable space-time actually has an orientation.

> If $< M, g, \nabla >$ is a temporally orientable space-time we can define in a globally consistent manner an equivalence relation $S(\, , \,)$ on the set of timelike tangent vectors which holds between two such vectors U and V just in case they have the same time sense. The quotient of the set of timelike vectors by $S(\, , \,)$ has two elements, O_1 and O_2. The choice of one element as containing the future pointing timelike vectors is the choice of a temporal orientation or direction of time for $< M, g, \nabla >$. (Earman 1974, p. 18)

The definition of a temporal orientation in turn allows to define an 'earlier' and 'later' relation.

[7] I will come back on the point that a certain asymmetry is generated by the choice of initial conditions in section 4.3.4 below.

[8] The following definitions are taken from Earman (1974, 2002), and any details that had to be left out here for the sake of brevity can be found in these articles.

> Let $<< M, g, \nabla >, O >$ be a temporally oriented space-time where O is the time orientation. Define the [asymmetric and transitive] relation $E(\, , \,)$ on $M \times M$, where $E(x,y)$ is interpreted to mean that x is chronologically earlier than y [and y is chronologically later than x.] (Earman 1974, p. 18)

It is now time to define time reversal invariance. Let T be a physical theory and $\mathfrak{M}_T^D$ be the set of all dynamically possible models, that is models that agree with the laws of T. "The time reversal transformation $\mathbf{T}$ is a bijective map $\mathbf{T}$: $\mathfrak{M}_T \rightarrow \mathfrak{M}_T$ such that $\mathbf{TT}(m) = m$ for any $m \,\varepsilon\, \mathfrak{M}_T$." (Earman 1974, p. 23, notation changed) Accordingly, a theory is time reversal invariant if the following holds:

> A consistent theory T is time reversal invariant just in case $\mathbf{T}(\mathfrak{M}_T^D){=}\mathfrak{M}_T^D$ (Earman 1974, p. 23)

It follows that, if model m has the temporal direction O, then the temporally reversed model $\mathbf{T}(m){=}m^{\mathrm{T}}$ will have the reversed temporal direction $\mathbf{T}(O){=}O^{\mathrm{T}}$, and if $\mathrm{E}(x,y)$ in m, then $\mathrm{E}(y,x)$ in m^{T}.

While the above gives a general definition of time symmetry, it is also important to have a look at how in general the transformation $\mathbf{T}$ is implemented in a theory. Let $S(t)$ be the state of a system in m, that is roughly the complete set of physical information about the system at time t. $\mathbf{T}(S(t)){=}S^{\mathrm{T}}(\text{-}t)$ is the temporally reversed state. As Earman (2002, p. 248) shows, $S{=}S^{\mathrm{T}}$ will only be true if $S(t){=}$const. This can be the case if the system consists entirely of stationary particles, however, in general a state and the temporally reversed state are different. For example, the time reversal operator $\mathbf{T}$ will reverse the velocities of particles or the orientation of the magnetic field. In other words, in general the time reversal operator does not only flip the sign of the time variable. Hence, if S_{i}, ..., S_{f} is a temporal evolution of a state S, then in general the time reversal ${S^{\mathrm{T}}}_{\mathrm{i}}$, ..., ${S^{\mathrm{T}}}_{\mathrm{f}}$ will be the temporal evolution of a different state S^{T}. We can put the definition of time reversal invariance into the following more informal definition:

> A theory is time-reversal invariant iff for any history S_{i}, ..., S_{f} allowed by the theory, the history ${S^{\mathrm{T}}}_{\mathrm{i}}$, ..., ${S^{\mathrm{T}}}_{\mathrm{f}}$ is also allowed by the theory, where S^{T} is the time reverse of S, as determined by the time-reversal operator. (Farr & Reutlinger 2013, p. 222, notation changed)[9]

With these definitions at hand, it is now time to ask whether the time reversal invariance of physics licenses **(¬TD)**, and indeed, there is an interpretation of time reversal invariance that seems to do that. This position has been dubbed the Reichenbach-Gold position by Earman (1974). One of the most telling sections for this position in Reichenbach's writings might be the following:

> The use of converse descriptions, that is, descriptions in negative time, is a convenient device for the study of time problems. Since it is always possible to construct a converse description, positive and negative time supply equivalent descriptions, and it would be meaningless to ask which of the two descriptions is true. (Reichenbach 1956, p. 31f.)

[9]But see also Albert (2000) for a different understanding of the time reversal operation.

Earman interprets this as a passive interpretation of time symmetry, according to which "a symmetry operation corresponds not to a change from one physical system to another but rather a change, so to speak, of the point of view from which the system is described". (Earman 1974, p. 26) I do not believe that this is the most plausible interpretation of Reichenbach, since, contrary to what Earman (1974, p. 23) reads, Reichenbach does not speak of two equal descriptions of the same world. Furthermore, Reichenbach (1956, p. 32) comes to the conclusion that even reversible processes can give rise to asymmetric, though not directed, causal connections, which seems to indicate that there is a difference between the cause coming before the effect and the reverse. Rather, from reading the context I would suggest that Reichenbach is merely stressing the point that from time reversal symmetry it follows that a given mechanical problem, for example finding the trajectory of a ball being thrown, can be solved in positive or negative time, that is, it is a convention whether we use t or $-t$ (cf. Sachs 1987, p. 5). However, for the sake of the argument one might interpret Reichenbach the 'Earmanian' way, and indeed Gold (1966) seems to promote a passive interpretation more explicitly.

(¬TD) follows directly from the passive interpretation. If A causes B and this relation has the temporal orientation O and the temporal precedence relation $\mathrm{E}(A,B)$ holds between A and B, then by the passive interpretation A and B also have the temporal orientation O^{T} and $\mathrm{E}(B,A)$. Hence, if A causes B and A comes before B, then it is also true that the same event B comes before the same event A.

What are the arguments or motivations for a passive interpretation of time reversal symmetry? Earman (2002, p. 255f.) suggests that the following was Reichenbach's motivation. The evolution of a system consisting entirely of particles can be described with the help of a betweeness relation $bet(p, q, r)$ that "can be taken as a primitive relation which does not presuppose a time orientation." (Earman 2002, p. 256) Even though betweeness does not presuppose time, it would nevertheless be possible to ad that either p comes before q and r or r comes before q and p. Still, both temporal directions would describe the same thing, that is, a system in which q lies between p and r. However, at least for the world we live in this does not seem to be an option, because "unlike pure particle mechanics, an electromagnetic world obeying Maxwell's equations is not happily described as a world that lacks a time orientation and is endowed with only a temporal betweenness relation." (Earman 2002, p. 256)[10]

Another motivation for the passive interpretation of time reversal symmetry might be the attempt to extend a passive interpretation of spatial symmetries. Given that it is plausible to interpret the group of spatial rotations and translations passively, it might be plausible to believe the same about temporal transformations. As

[10]Earman (1974, p. 27-29) discusses such a reductionism as a motivation for the passive interpretation in more general terms and at greater length. Thus, reductionism about time could lead to the following argument: 1. Any relation of temporal precedence E(,) must be reduced to a non-temporal relation R(,). 2. There is no non-temporal relation R(,) and no non-temporal property that can distinguish between two opposite temporal directions. 3. Therefore, there is no direction of time. As Earman explains, there is currently no reason to believe that such an argument is true.

Earman (1974, p. 23) points out, a consequence of this suggestion is that either the transformations of charge conjugation **C** and parity **P** have to be interpreted passively as well, or there has to be some difference between **C** and **P** on the one and **T** on the other side. Since there is no proposal for what could establish such a difference, only the first possibility is left. To treat **C**, **P** and **T** in the same way, I believe, is also strongly entailed by the fact that not all of today's physics is invariant under **T**. As is well known, the decay of the K-meson is not invariant under **T**, but only under the combined transformation **CPT**. Hence, if under this condition it should still be true that physics makes no difference between past and future, **C** and **P** must be interpreted passively as well. Otherwise the **CPT** transformation would relate different processes, different in charge and parity, and it could hardly be argued that, from the passive interpretation of **T**, a **CPT** transformed process is the same process observed from a different perspective.[11]

Having said that, the proponent of the passive interpretation is not in a better position now, as Earman points out:

> But many physicists regard this and similar conclusions about **C** and **P** as a reductio ad absurdum of the passive interpretation of the discrete symmetries **C**, **P**, and **T**. The passive interpretation of continuous symmetries, like spatial rotation, is meaningful since one can suppose at least in principle that an idealized observer can rotate himself in space in correspondence with the given spatial rotation. [...] But how is an observer, even an idealized one, supposed to 'rotate himself in time? (Earman 1974, p. 27)

Earman here highlights an important difference between continuous and discrete symmetries,[12] which however is denied by Price, who in turn endorses the passive interpretation of **T**:

> Again, I think Earman is quite correct to treat **C**, **P**, and **T** as on a par here. But the challenge that follows seems unsuccessful. The obvious reply is that the passive interpretation of the symmetries requires not the rotation of the original idealized observer, but only (at most), the idealized postulation of a second observer who stands in the appropriate 'rotated' relationship to the first observer. Far from being unimaginable in the temporal case, this is exactly the possibility that Boltzmann imagines – and for real observers, too, not merely idealized ones. (Price 2011, p. 289)

[11] Atkinson (2006, p. 534) puts even more weight on the asymmetry of the K-meson decay: "Nevertheless, one might well make a point of principle: because of the breaking of T symmetry by the weak interaction, it is not true that there is no microscopic arrow of time in electrodynamics. Whether or not there is more to be said concerning the origin of a macroscopic arrow, the mere whiff of directionality at the level of the fundamental laws calls into question the claim that the distinction between past and future is merely one of our standpoint as agents (Price, 1996, p. 168)." Having said that, Aitkinson does not think that there is an arrow of time in quantum electrodynamics, because it is symmetric under **CPT** transformations.

[12] The difference between continuous spatial symmetries and **T** symmetry might also be supported by the fact that only the former, but not the latter lead to conservation laws (cf. Sachs 1987, p. 6).

I will have more to say on Boltzmann's view that Price mentions here in due course. Before, I wish to emphasise that Price's reply seems to miss the point entirely, since the question is not whether a transformed observer is possible, but whether this observer is still looking at the same system. Why is it that in the case of continuous transformations we believe that an observer who 'transforms' himself still looks at the same system, and therefore a passive interpretation seems feasible? Because we can look at the same thing from different angles or different distances. If I turn my coffee mug around or place a bit further away from me, it is save to say that I am still looking at the same coffee mug. We cannot, however, transform ourselves in any way to see the mirror image of something or anti-particles instead of particles. Earman does not need to deny that it is possible to postulate an observer, who lives in a mirror world, a world made from anti-particles or a temporally reversed world; such worlds, including observers, are possible. However, Price must give an additional argument to the effect that the transformed observer is still in the same world and not in a different one. If I look at my coffee mug and notice that today it is the mirror image of the mug that I had yesterday, then most likely I would suppose that I am looking at a different mug, and not that somehow I have changed my perspective of looking at the same mug like yesterday. A physicist, conducting an experiment on electrons, can 'transform' himself as much as he likes, he will not see positrons instead of electrons. Likewise, it is one thing to postulate a temporally rotated observer, and quite another to rotate oneself in time. Again, to put the argument in different words, the transformations **C**, **P** and **T** make a difference in the theoretical description of a system that changes its dynamics and therefore is also an empirical difference. Both cannot be said for continuous transformations like rotations. The promoter of the passive interpretation has to give a reason why these differences only *appear* as differences, but are not real, and such a reason does not seem to be forthcoming.[13] As a result, it seems more reasonable to interpret the theoretical and empirical differences as real differences.[14]

The conclusion of this discussion is that the time reversal symmetry of physics is best not interpreted passively and therefore the passive interpretation cannot support **(¬TD)**. Nevertheless, there might be a different route on which the time reversal symmetry of physics can lead to **(¬TD)**, which I briefly want to discuss. Price (1996b) champions what he calls the atemporal view and within his book one frequently finds statements that seemingly come very close to the Reichenbach-Gold

[13]On the contrary, here is a problem that I believe has to be answered by the passive interpretation. The dynamics determine the past and future history of a system; different dynamics, different history. If the passive interpretation was true and one and the same system could have different dynamics, then this would amount to a breakdown of determinism (similar to the hole argument in the general theory of relativity). A given state of a system would not determine the past and future histories, since at any point of time we are free to change the dynamics by applying a transformation.

[14]In yet another argument Earman (1974, p. 24) considers that from the Reichenbach-Gold view it follows that possible worlds that are different with respect to temporal orientation collapse into one world, if there is no other feature that can distinguish them. Earman obviously regards this to be a drawback, while Price (2011, sec. 3.7) seems to be less worried.

position. For example:

> Again, this follows immediately from the realization that there is nothing objective about the temporal orientation. A universe that collapses without deflation just *is* a universe that expands without inflation. It is exactly the same universe, under a different but equally valid description. (Price 1996b, p. 86)

A statement like this seem to imply that the same system can equally be described by using opposite temporal orientations, and thus would vindicate **(¬TD)** similar to the passive interpretation (cf. Savitt 1996, p. 364). To see whether this is the case, it will be necessary to have a closer look on what Price means. Price (1996b, ch. 4) argues in favour of a cosmological picture that he calls a Gold universe. A Gold universe has two low entropy states and in between them a high entropy state. It starts to evolve from a low entropy state into a high entropy state, where the arrow of entropy flips around, so that entropy decreases again. Assuming that entropy correlates with size, the Gold universe will be small in the low entropy state and large in the high entropy state. Price's main argument for the Gold universe is his 'basic dilemma', that is, "we seem to be presented with a choice between Gold's view, on the one hand, and the conclusion that the smooth big bang is inexplicable (at least by a time-symmetric physics), on the other." (Price 1996b, p. 82) The latter possibility is unacceptable for Price and therefore we are left with the first one. The need to posit low entropy states at both ends comes from the time reversal symmetry of physics, since positing a low entropy state only at one end would be a temporally asymmetric assumption that according to Price is incompatible with temporally symmetric physics.

It can reasonably be doubted that Price's argument is sound, in particular whether positing only a low entropy past would need an explanation,[15] however, this would require a lengthy discussion that I cannot give here. More important at the present point is whether a Gold universe supports **(¬TD)**. For the following reason I think that it does not. Let us call the low entropy end points of the Gold universe events A and C, while the high entropy state in between is B. If the atemporal view is correct, then it might follow from time reversal symmetry, that A and C can both be either the initial or the final state of the universe. Accordingly, if O_1 and O_2 are the possible time directions, then it is a matter of free choice, whether we assign O_1 to the phase between A and B and O_2 to the phase between B and C or vise versa. Thus Price (1996b, p. 84) is right in concluding "that there is no objective sense in which this reverse way of viewing the universe is any less valid than the usual way of viewing it. Nothing in physics tells us that there is a wrong or a right way to choose the orientation of the temporal coordinates." However, once this choice is made and, e.g., O_1 is assigned to the phase between A and B, then O_2 is not the temporal direction for the phase between A and B. If X and Y are two events between A and B and $\mathrm{E}(X,Y)$, then $\mathrm{E}(Y,X)$ will not be true. Even in an extreme case where the phase between B and C is the exact time reversed copy of the phase between

[15]See Callender (1998, 2004) and the response in Price (2004).

A and B, like a movie running backwards, events in the two phases will be at last numerically distinct. It is therefore not the case that in a Gold universe, if A comes before B, then the same event B also comes before A, and **(¬TD)** does not follow.

Without going into detail, I content that an analogous argument can be made in the case of a Boltzmann universe, which has for example been considered by Price (2011). According to Price, it was Boltzmann's view that in the universe there might be regions of space-time with opposite temporal directions.[16] Again time reversal symmetry might imply that there is no objective fact as to which of the opposite directions points into past or future. However, again it does not follow that opposite directions are true of one and the same region at the same time. Accordingly, Price comes to the following conclusion:

> Would Boltzmann deny that there is a difference between asteroid moving from the vicinity of Earth to that of Mars and the same asteroid moving from the vicinity of Mars to that of Earth? I think we can be sure that he would not! Rather, he would maintain that while of course the two cases are *different*, there's no fact of the matter as to which is which. (Price 2011, p. 283)

I do not wish to comment further on whether time has an objective direction or Price's atemporal view is correct. For the moment I am content with the conclusion that even without an objective direction of time, time reversal symmetry does not imply **(¬TD)**. Earman brings this to the point when he concludes:

> Reversibility for a physical system or invariance under time reversal for laws governing that system does mean, intuitively speaking, that motion makes good sense either way, from A to B or from B to A. But it does not mean that there is no distinction between these two senses or that the distinction is only verbal as Reichenbach and Gold would have it. (Earman 1974, p. 33)[17]

Having said that, from the failure of the directionality argument based on time reversal symmetry alone it does of course not follow that physics is sufficiently asymmetric to ground the temporal asymmetry of causation. I will return to this point in section 4.3.1, but before I wish to present a more successful version of the directionality argument.

4.2.3 The causal directionality argument

As Farr & Reutlinger (2013) point out, the directionality argument based on time reversal symmetry is not the only argument on the incompatibility of causation and physics that can be found in Russell (1913). In the following passage, Russell is making a directionality argument based on a feature of physics different from time reversal symmetry:

[16] According to Zeh (2006) this is possible, though very unlikely.

[17] In a more direct fashion, Mirman (1975, p. 496) claims that there "is no relation between the existence of a direction of time and the existence of time-reversal invariance." From this perspective, the failure of the directionality argument from time reversal symmetry can hardly be surprising.

> In the motions of mutually gravitating bodies, there is nothing that can be called a cause, and 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 configuration and velocities at one instant, or the configurations at two instants, render the configuration at any other earlier or later instant theoretically calculable. That is to say, the configuration at any instant is a function of that instant and the configurations at two given instants. This statement holds throughout physics, and not only in the special case of gravitation. But there is nothing that could be properly called 'cause' and nothing that could be properly called 'effect' in such a system. (Russell 1913, p. 14)

What Russell is saying is that the laws of physics exhibit determinism in both temporal directions, according to his definition of determinism. That is, given a system in state S_1 at time t_1, the state of the system at any other time can be calculated from a function of S_1, i.e., $S(t) = f(S_1)$. Even though this might not be the best definition of determinism (cf. Earman 1986, sec. II.5), more important at the moment is that it is not obvious at least, where the conflict with causation lies. Given that Earman (1986, sec. II.2) is correct in claiming that determinism and causation are two concepts that should best be kept separate,[18] there is no direct contradiction between the fact that physics is deterministic in both temporal directions and at the same time describes causal connections. However, following Farr & Reutlinger (2013) I believe that the conflict can be established, if Russell's statement is read as referring to bidirectional nomic dependence of physics. It is bidirectional nomic dependence that establishes the connection between Russell's determinism and causation.

According to Lewis nomic dependence can be defined as follows:

> The family C_1, C_2, ... of propositions depends nomically on the family A_1, A_2, ... iff there are a nonempty set $\mathfrak{L}$, of true law-propositions and a set $\mathfrak{F}$ of true propositions of particular fact such that $\mathfrak{L}$ and $\mathfrak{F}$ jointly imply (but $\mathfrak{F}$ alone does not imply) all the material conditionals $A_1 \rightarrow C_1$, $A_2 \rightarrow C_2$, ... between the corresponding propositions in the two families. (Lewis 1986b, p. 167)

Given this analysis, bidirectional nomic dependence of physics follows automatically from Russell's determinism.[19]

(PND) *Physics and nomic dependence*: For a deterministic system it is a law L that $S(t) = f(S_i)$, $i = 1, 2, ...$, and given the right conditions F, and S_1 and S_2 are two states of the system at times $t_1 < t_2$, then it follows that $S_1 = f(S_2)$ and $S_2 = f(S_1)$, and therefore $S_2 \rightarrow S_1$ and $S_1 \rightarrow S_2$, that is S_1 and S_2 are mutually nomically dependent on each other. A deterministic system therefore exhibits nomic dependence in both temporal directions, under the right conditions F.

[18] This is not meant to imply that causation and determinism are incompatible. See the discussion of Mackie (1980), who argues for incompatibility, in Earman (1976, p. 12).

[19] Without going into detail, I content that a similar analysis can be made for indeterministic systems (cf. Farr & Reutlinger 2013).

It is important to keep in mind that time reversal symmetry and bidirectional nomic dependence are two different features of a theory. As Farr & Reutlinger (2013, p. 227) conclude: "The key point we wish to stress is that, although it may be the case that a nomically unidirectional theory must also be time-reversal non-invariant, the converse is not true: the time-reversal non-invariance of a theory *does not* entail that the theory is nomically unidirectional." Hence, the directionality argument from time reversal symmetry and the argument developed in this section are two different arguments that require different solutions.[20]

The connection between causation and nomic dependence, however, is less obvious. For Farr & Reutlinger, it is established by a realist reading of nomic dependence: "According to this realist reading, nomic dependencies correspond to ontic dependencies holding between token states of the actual world, and as such, bidirectional nomic dependence is directly comparable to the features of causation with which the Directionality Argument is concerned." (Farr & Reutlinger 2013, 227) This realist reading is a necessary condition for constructing the argument against a causal interpretation of physics. Of course, not everyone, especially Humeans, has to agree with such a realist reading. It might be the case, for example, that all relations that exist are spatiotemporal relations. For the sake of the argument, I take the realist reading as given.[21]

With this presupposition, the directedness of causation then has to correspond to a directedness of nomic dependency. Otherwise it would be the case that a directed relation, causation, corresponds to a symmetric relation, nomic dependence, which would be unclear and against the realist reading of nomic dependence. This is captured in the following principle:

(CND) *Causation and nomic dependence*: If A causes B then B nomically depends on A, and if A causes B then A does not nomically depend on B.

In what follows, I assume that **(CND)** is a necessary condition for a causal interpretation of physics.

On the basis of the foregoing discussion, the following directionality argument from bidirectional nomic dependence can be made (cf. Farr & Reutlinger 2013, p. 229):

1. If a system is causal, then, by **(CND)**, it does not exhibit bidirectional nomic dependence.

2. If a system is physical, then, by **(PND)**, it exhibits bidirectional nomic dependence.

[20] It is worth adding that time reversal symmetry implies bidirectional nomic dependence, in so far a theory is deterministic at all. However, bidirectional nomic dependence does not imply time reversal symmetry (cf. Earman 2002, p. 254).

[21] However, as Farr & Reutlinger (2013, 228) highlight, Humeanism and a realist reading of nomic dependence might be compatible, if nomic dependence is seen as a non-fundamental relation. In consequence, Farr & Reutlinger remain neutral with respect to Humeanism.

3. If a system is physical, then it is not causal.[22]

To end this section, I briefly want to discuss a response to this kind of argument by Ney (2009). Ney's reaction consists not in attacking the directionality argument, but in the suggestion that causation need not be an asymmetric theory.[23] According to Ney (2009, p. 752) the asymmetry of causation stems from the more anthropomorphic notions of "explanation, prediction, and action". Thus, a theory of causation at the fundamental physical level, deprived of unnecessary anthropomorphic notions, is not necessarily asymmetric. In particular, Ney (2009, p. 753) claims that even without asymmetry there "is still causation, because there is still physical determination." However, Farr & Reutlinger (2013, 233) have replied that the confrontation between Russell and Ney "looks like a merely verbal dispute. The parties in the dispute are not really disagreeing about anything." In particular, they agree on that causation usually is understood as an asymmetric relation and on the validity of the directionality arguments. Only where Russell opts for the abolishment of causation, Ney pleads for a shift in concept. As a consequence, Farr & Reutlinger suggest that it is best to not follow Ney, since in the current debate 'causation' is understood as an asymmetric relation, and changing the meaning of the concept will not solve any of the problems around asymmetry, but more likely will only cause confusion. A further problem may arise from Ney's suggestion to separate causation from explanation. It is not entirely clear what she means here, but given the widely shared view that there are at least some causal explanations, its seems difficult to argue that causation has nothing to do with explanation at all. Hence, Ney might best be interpreted as saying that explanation is not identical or cannot be reduced to causation. In that case, one has three options: First, a top-down approach in which we first identify explanatory relations and on that basis identify causal relations. Second, a bottom-up approach, which takes the reversed direction (cf. Psillos 2002, 282-83). Third, a combination of the two preceding option. For all three options it is problematic to have symmetry at one end, but asymmetry at the other; neither can symmetry be deduced from asymmetry nor the other way round. Ney's proposal thus leads to the highly questionable consequence that if causation is symmetric, then explanations have to be symmetric as well, which might be regarded as a reductio ad absurdum.

In conclusion, I believe that the directionality argument from bidirectional nomic dependence poses a much bigger challenge for the causal interpretation of physics than the analogous argument from time reversal symmetry. Having said that, I will spend sections 4.3.2 and 4.3.4 arguing that it can be refuted, too.

[22]The argument is based on a deterministic theory. However, as Field (2003, p. 437) points out, it works equally well for a theory that is indeterministic in both temporal directions.

[23]See also Blondeau & Ghins (2012) for an account of symmetric causation.

4.3 The compatibility of causation and physics

4.3.1 Initial conditions, the temporal asymmetry of the facts and local energy flux

The result of section 4.2.2 was that the received directionality argument, supposed to show the incompatibility of time reversal symmetry and causation, fails. This by itself does not license the conclusion that physics is sufficiently asymmetric to allow for a causal interpretation, as it has not been shown that time in fact has a direction. The question whether time has a direction is probably tantamount to one of the most convoluted bundles of problems in the philosophy of physics. If there should be hope to make any progress within the limited boundaries of the present work, it needs to be exactly clear what has to be shown and what not. Hence, to begin with I wish to argue that two of the most controversial questions about time do not have to be answered given the aims of this work. These questions are whether time has a single global direction, and whether there is an objective fact which time direction is future and which one is past.

Several different proposals as to what could give time a single global direction have been given. To name only the most popular, a global direction of time has been derived from certain initial conditions of the universe together with statistical mechanics, or the direction of electromagnetic radiation possibly together with special absorber configurations, or certain possible solutions to the quantum measurement problem. As is well known, none of these ideas has lead to any overwhelming agreement among philosophers. But is an answer to this question necessary to show that processes in QFT can be interpreted as causal, which after all is the goal of this work? QFT is the description of scattering or decay processes of particles that happen on very small time and length scales. The discovery of a global direction of time could presumably also provide a single time direction for processes in QFT. However, the failure to find a global direction of time does not imply that processes in QFT are not temporally directed. If processes in QFT are temporally directed, then there are two possibilities: either all processes have the same direction, or the direction can vary from process to process. In the first case, the grounding of the temporal asymmetry of causation in QFT should be unproblematic, and also for the second case I have already argued in section 4.2.2 that even a universe with different temporal directions can be interpreted causally. The bottom line is that for a causal interpretation of QFT it should be sufficient that individual processes in QFT have a temporal direction, but not necessary that all processes have the same temporal direction. This seems to be the minimal requirement for a causal interpretation that is easier to defend than a global direction of time, and this is what I will argue for in what follows.[24]

The second question that I will not answer is whether there is an objective fact

[24] My account therefore stands in contrast to Albert (2000) and Kutach (2007), who both ground the asymmetry of macroscopic causal processes in a combination of thermodynamics and the past hypothesis (cf. Frisch 2005 for a critique of this approach).

which time direction is future and which one is past. This is, of course, not a terminological question about which direction of time the term 'future' or 'past' should be assigned to, but rather the question whether time itself is different, i.e., anisotropic in both directions. This question is more opaque than the previous one, as it is not entirely clear what temporal anisotropy means and what could provide it. This topic would require a lengthy discussion that cannot be given here. Thus, I wish to simply stick to the definition in Earman (1974, p. 29) according to which:

> A temporally orientable space-time $< M, g, \nabla >$ is literally temporally isotropic, the same in both time directions, if there is a diffeomorphism d of M onto itself which reverses the temporal orientations but which preserves g and ∇ (i.e. d is an isometry of g and therefore an affine mapping of ∇).

Accordingly a spacetime is temporally anisotropic if this condition is not fulfilled. The important point is that under this definition neither temporal isotropy nor anisotropy is a necessary condition for time to have an orientation. Both possibilities are compatible with there being a vector field on M that gives the direction of time.[25] Thus, we do not need to know whether time is isotropic or not, in order to know whether time has an orientation. Similarly, I want to suggest that neither temporal isotropy nor anisotropy is a necessary condition for the causal interpretation. If it should turn out that time has a direction O according to the definition given in section 4.2.2, then it can be said that causation is temporally asymmetric with respect to O, without any implication on isotropy. This seems plausible since, at least outside the realm of general relativity, causation is a feature of processes in time, which puts it closer to an asymmetry given by a vector field in space-time, rather than an asymmetry of space-time itself. Defining the direction of time through a vector field does not mean that the time is identified as the vector field. Similarly, causation is not a temporal relation, but the temporal asymmetry of causation means that causal relata are related in a special temporal way. In conclusion, it can be said that for a causal interpretation of physics it is necessary that there is an objective distinction between past and future, but not that time itself is objectively different in past and future. What has to be shown in this section therefore boils down to that individual processes in QFT have a temporal direction O.

As is well known, our world exhibits several different temporal asymmetries, that is, processes that only occur in one temporal direction, but not the other. This might be surprising (cf. e.g. Mirman 1975, p. 493), given the time reversal symmetry of physics, however, this 'asymmetry of the facts' is not in conflict with the symmetry of the laws. As is also well known, time symmetric laws are compatible with asymmetric boundary conditions:

> It is essential to keep in mind that time-symmetric laws are perfectly compatible with asymmetric solutions. Almost all solutions of the fundamental equations of motion are time-asymmetric [...] The symmetry of the laws of motion requires only that for every asymmetric solution that is realized in nature there

[25] Cf. Price (2011, p. 292): "The contents of time – i.e., the arrangement of physical stuff – might be temporally asymmetric, without time itself having any asymmetry."

> must mathematically – not necessarily physically – exist precisely another, time-reversed one. (Zeh 2012, p. 208)[26]

Despite time symmetric laws, processes do not have to look the same in past and future; or in other words, the initial state of a process can be different from its final state. As Earman (2011, p. 489) points out, this kind of asymmetry can be interpreted as having either "a de facto or a law like character". It seems to me that choosing the latter possibility would mostly be motivated by an attempt to explain a global direction of time. Furthermore, to promote the asymmetries to law like facts should be much harder to make plausible than to opt for the other possibility and simply content that they are contingent facts. For the present work nothing hinges on a decision between these possibilities. For the causal interpretation it is sufficient that actual processes are asymmetric, and I will therefore continue to speak of the asymmetry of the facts, be they de facto or law like.

A particular and much discussed factual asymmetry is that of electromagnetic radiation, which I want to look at first before going into QFT. A general solution to the classical Maxwell equation can be written as

$$F^{\mu\nu}(x) = \lambda \left(F^{\mu\nu}_{ret}(x) + F^{\mu\nu}_{in}(x)\right) + (1-\lambda)\left(F^{\mu\nu}_{adv}(x) + F^{\mu\nu}_{out}(x)\right), \quad 0 \leq \lambda \leq 1$$

with $F^{\mu\nu}$ as the relativistic field strength tensor. $F^{\mu\nu}_{ret}$ and $F^{\mu\nu}_{adv}$ are the solution to the inhomogeneous Maxwell equations with a source, while $F^{\mu\nu}_{in}$ and $F^{\mu\nu}_{out}$ solve the homogeneous Maxwell equations without a source (cf. Earman 2011, p. 492). Furthermore, $F^{\mu\nu}_{ret}$ and $F^{\mu\nu}_{in}$ lie in the past light cone of t, while $F^{\mu\nu}_{adv}$ and $F^{\mu\nu}_{out}$ lie in the future light cone of t. This general solution is temporally symmetric (cf. Zeh 2007, p. 21 f.), because $F^{\mu\nu}_{ret}(x) + F^{\mu\nu}_{in}(x) = F^{\mu\nu}_{adv}(x) + F^{\mu\nu}_{out}(x)$, however, processes observed in our world are not described by the general solution. Everything we observe are retarded fields $F^{\mu\nu}_{ret}$, and as Earman (2011, p. 492) points out, retarded and advanced solution "are generally *different solutions* of the inhomogeneous Maxwell equations – rather special time symmetric cases in which these fields are equal are the exceptions that prove the rule."[27] In the literature, it has been discussed at great length how this asymmetry can be explained and whether it is fact or law like.[28] For present

[26]Cf. Callender (1997, p. 227), Sachs (1987, p. 2 f.) and Earman (2002, p. 254) for similar statements.

[27]See also Frisch (2000, ch. 3), arguing against Price (1996b) that there is a difference between the absorption and emission of radiation, that is, why emission is not simply the time reverse of absorption and advanced are not the same as retarded fields.

[28]Rohrlich (2000, p. 1) grounds the asymmetry of radiation in the asymmetry of causation: "Since causality (the time order of cause and effect) requires *retarded* rather than advanced self-interaction, it is causality which is ultimately responsible for the arrow of time." I do not endorse this claim here, since, justifying a physical assumption by a philosophical concept is the opposite of the method that I follow.

Price (1996b) and Zeh (2007) on the other hand offer different explanations based on the Wheeler-Feynman absorber theory.

purposes there is no need to enter these discussions, rather I will merely take the fact that only retarded fields are observed as the basis for the following argument.[29]

The field strength tensor can be used to define a temporal direction O as follows. First it can be used to define the energy-momentum tensor:

$$T^{\mu\nu} = \frac{1}{4\pi}\left(F^{\nu}_{\alpha}F^{\mu\alpha} + \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}\right)$$

of which the component T^{0i}, $i = 1, 2, 3$, is proportional to the Pointing vector and can be interpreted as the flow of energy. Earman now describes how on this basis a direction of time, that satisfies the definition given in section 4.2.2, can be defined:

> Let U^α be a continuous non-vanishing timelike vector field. Then the four-momentum $W^\alpha := -T^{\alpha\beta}U_\beta$ as measured relative to U^α is a continuous timelike field. If U^α and W^α point in the same direction at some point where $F^{\alpha\beta}$ is non-zero, they will point in the same direction at every point where $F^{\alpha\beta}$ is non-zero. If they do point in the same direction, U_α defines the future pointing (relative to $F^{\alpha\beta}$) time orientation. (Earman 2002, p. 256; notation changed)

However, if, contrary to what Earman seems to have in mind, it is not required that there is a single global time direction, then this definition seems too strong. As Zeh (2007, p. 29 f.) shows, the retarded field $T^{\alpha\beta}$ has its support in the future light cone. Therefore, the direction of the electromagnetic field can be used to distinguish between past and future light cones, without a global time direction. In particular, the energy flow, given by the Pointing vector, of a retarded field will be directed and can be said to be from the past to the future. (cf. Castagnino & Lombardi 2009, p. 16 f.; Wohlfarth 2013, p, 85 f.)

According to Castagnino & Lombardi (2009) the above definition of the direction of time extends to QFT. As they point out, one of the Wightman axioms is that the "spectrum of the energy-momentum operators P^μ is confined to the (closed) forward cone $p^2 \geq 0$; $p^0 \geq 0$." (Haag 1996, p. 56) Thus any measured energy-momentum p^μ in QFT is temporally asymmetric and, by definition, pointing in the future direction.[30] Furthermore, this means that any energy-momentum transfer in QFT goes from the initial to the final state (cf. Aiello et al. 2007, p. 31). The inclusion of this constraint into the axioms of a mathematically rigorous version of QFT might be interpreted as giving the temporal direction a law like status. However, I do not wish to go that far and merely take it as making clear the fact that any known process in QFT today agrees with this temporally asymmetric constraint. In other words, I do not put any argumentative weight on the *axiomatic* status. This is also required by the fact that the energy condition does not have the

[29] Thus, without denying that an explanation of this facts might be possible, my position is what Frisch (2000, p. 405) calls the 'textbook account' "that offers little more than the statement that the asymmetry does in fact hold." However, contrary to Frisch I also do not want to claim that the retardation condition is another law of nature.

[30] For Heathcote (1989, p. 95) this axiom is already sufficient to claim that "there is no backward causation" in QFT.

status of an axiom in non-axiomatic QFT. Nonetheless, as Aiello et al. (2007, sec. IV. B) explain, textbook QFT features it as an empirically motivated constraint in the following way. The group of proper Lorentz transformation, under which four-momenta in QFT transform, has two discrete components. The first is the orthochronus component, which preserves the direction of time, and the second is the non-orthochrous component, which changes the direction of time. It is standard textbook procedure to dismiss the non-orthochronus component as not physical (cf. e.g. Maggiore 2005, p. 17 and Duncan 2012, p. 109). Together with the choice of the metric signature $\eta_{\mu\nu} = (+, -, -, -)$, this implies again the condition that $P^{\mu}P_{\nu} = p^2 \geq 0$. Castagnino & Lombardi (2009, p. 24) sum up this result, when they write:

> In turn, the local energy flow is what provides the substantial criterion for choosing one of the time-symmetric twins. In consequence, by endowing the arrow of time with a local physical meaning, the local energy flow can be used for breaking the symmetry of the set of solutions resulting from the time-reversal invariant laws of local theories.

How does this help for the causal interpretation of physics? First of all, it shows that processes in QFT are sufficiently asymmetric to define an objective distinction between past and future. Furthermore, if causation can be tied to the transfer of energy-momentum, then this would make causation automatically directed in time. Doing so, however, requires a longer argument that I will give in chapter 7. Nevertheless, assuming for the sake of the argument, that causation can be reduced to the transfer of energy-momentum, I wish to point out that thereby it is not a matter of definition that whatever comes first in time is the cause and what is later is the effect. Hence, the problems associated with such a position, presented in section 3.3.5, are avoided. Rather, the cause lies at the origin of the vector of energy-momentum transfer, the initial state, while the effect lies where the vector points at, the final state.[31]

Having said that, one might object that it is a convention in physics what is called past and what future. What happens if physicists conventionally decide to interchange these terms, does it follow that now the cause lies in the future and the effect in the past, and does this contradict **(TD)**. I do not think that this poses a problem. It has to be kept in mind that **(TD)** is merely the articulation of a commonly shared intuition and not a statement of an a priori truth that would tie the terms 'cause' and 'past' together. As much as it is a convention what in a Minkowski diagram is called past and what future, it is a convention that causes lie in the past and effects in the future. If the folk should decide to reverse this convention then **(TD)*** would be true according to which the effect comes before the cause. The agreement of **(TD)** with the terminology in physics merely rests on the agreement of two conventions. In turn, a disagreement between these conventions does not pose a problem, since nothing substantial rests on conventions. In contrast, what is not

[31] Price & Weslake (2009) argue against 'third arrow strategies', but do not discuss the possibility that I suggest here.

a convention is that there is a difference between where energy-momentum comes from and where it flows to. Given that this flow defines the direction of time and that one of the endpoints of the flow represents the cause and the other the effect, causation has a temporal direction that goes beyond mere convention.

4.3.2 How interventions do *not* solve the problem

The temporal asymmetry of the facts will be central for the argument how causation and the nomic bidirectionality of physics can be reconciled with the help of interventionism. Before presenting this argument in the next sections, I wish to briefly consider previous, and to my mind unsuccessful, attempts to base a similar argument on interventions. That causal actions or interventions bring with them certain asymmetries has already be noted by a number of authors. Interventions seem to be temporally asymmetric in that we can only intervene on the future. And they seem to be causally asymmetric in that there is a difference between doing something and the result of that action. It is therefore no surprise that it has been tried to ground the asymmetry of causation on the asymmetry of intervention and thereby solve the two prominent problems that have been presented in the last sections. In this section I will discuss these solutions and give reasons for why I believe they are unsuccessful.

I have already criticised the agency theory of causation by Menzies & Price in section 3.3.4 as having no implications for a theory of physical causation and will not repeat this critique here. However, in a newer version of the agency theory, Price & Weslake (2009) claim to solve the problem of the temporal asymmetry of causation and it is worth to have a quick look on their proposal. According to them, the temporal asymmetry of causation can be reduced to the temporal asymmetry of deliberation, viz., the fact that to advance one's goals it only makes sense to deliberate towards the future, but not the past. Omitting many details, their account of deliberation can be summarised as a counterfactual theory of deliberation, that is, when deliberating what to do, we have to evaluate the truth values of certain counterfactuals. This evaluation in turn, as Price & Weslake explain, crucially depends on the condition 'holding the past fixed'. That is, to evaluate the consequences of one's action, one must assume that nothing in the past of the action changes. A desideratum of this condition is the perspective of the deliberator:

> If the relevant species of counterfactual reasoning develops from the kind of hypothetical reasoning needed in epistemic deliberation, the principle that one should hold the past fixed provides a simple codification of the asymmetry of the deliberator's perspective [...]. (Price & Weslake 2009, p. 437)

Furthermore, since the aim of Price & Weslake was to explain the temporal asymmetry of causation by reducing it to the asymmetry of deliberation, it must be concluded that 'holding the past fixed' is what provides the asymmetry of causation as well.

However, already Reichenbach (1956) has put forward a criticism of the attempt to explain the asymmetry of causation by the asymmetry of intervention, which I believe is also applicable for the asymmetry of deliberation. According to Reichenbach, to explicate the meaning of a claim like 'We can only intervene on the future but not the past', it is necessary to make counterfactual claims about what would have happened, if we had not intervened. And counterfactuals of the form 'If intervention I had not happened, event A would not have happened' presuppose that the past remains unchanged, or else there could be another event B that causes A, making the counterfactual false. Reichenbach (1956, p. 45) concludes:

> No wonder that acts of intervention change only the future, and do not change the past; the term 'intervention' is defined by the condition that the past be unchanged. The statement that acts of intervention cannot change the past is a trivial tautology. This consideration leads to the conclusion that acts of intervention cannot define a direction of time. The term 'intervention' is defined only after a direction of time is given; acts of intervention are actions that leave the past unchanged.

I content that this criticism also applies to Price & Weslake. Their reduction of the asymmetry of causation to the asymmetry of deliberation merely shifts the problem and does nothing to explain or ground the asymmetry of causation; the same questions that can be asked about causation can be asked about deliberation, for example, whether the fact that we only deliberate towards the future can be grounded in physics. It might be that these questions are easier to answer in the case of deliberation than in the case of causation, but Price & Weslake do not give any reasons to think so. The lesson of this discussion is that the simple fact that we can only deliberate or intervene towards the future does not provide a sufficient explanation for the temporal asymmetry of causation.[32]

Similar to what has been discussed above, Frisch (2010) makes an argument for temporally asymmetric causal relation in physics based on the fact that we can only intervene on the future. However, Frisch analyses the asymmetry of intervention as a particular asymmetry of state preparation in physics, thereby supposedly giving it more substance than the mere tautology has, which Reichenbach was criticising. Frisch's argument is the following: 1. There is a temporal asymmetry of state preparation in physics. 2. The best explanation for this asymmetry is that it is an asymmetry of causal relations. 3. Therefore, by inference to the best explanation, there are asymmetric causal relations in physics. But what does Frisch mean by the asymmetry of state preparation? Here is how Frisch (2010, p. 81) explains the asymmetry:

> Consider a system S that is governed by both past and future deterministic laws. That is, let us assume that the final state $S_i(t_f)$ of the system is

[32]Another noteworthy account is found in Kutach (2011), who is not building up on any asymmetry of intervention per se, but only on an asymmetry of 'useful intervention'. I believe it is obvious that such pragmatic asymmetry does not help in clarifying issues about whether physical processes are causal. Furthermore, Kutach equates intervention (he uses the maybe more general term 'influence') with the temporally symmetric notion of nomic dependence, and acknowledges that as a consequence there are causal processes in both temporal directions. In effect, suggesting a temporally symmetric variety of causation sets Kutach close to Ney (2009), who I have already criticised in section 4.2.3.

> uniquely determined by the initial state $S_i(t_i)$, where $t_i < t_f$, together with the dynamical laws and the boundary conditions; and that the initial state $S_i(t_i)$ is similarly determined by the final state $S_f(t_f)$. Thus, if S is closed between t_i and t_f, , then the initial and final states are both dependent on each other. Nevertheless there is an asymmetry of state preparation in the following sense. We can prepare the system in its initial state $S_i(t_i)$ without making use of any knowledge we might have of the system's dynamical evolution between t_i and t_f; and we can subsequently calculate the system's future evolution for times $t > t_i$ from the initial state, the dynamical laws, and the boundary conditions. But we could not similarly first prepare the system's final state at t_f without using our knowledge of the dynamics and then take the final state together with the laws to calculate the system's past evolution for $t < t_f$. (Of course we cannot first prepare the system in S_f and then let it *evolve* into S_i. That is not what the asymmetry consists in. Rather the asymmetry consists in the fact that we cannot first prepare the system in S_f without *making use* of facts about the dynamical evolution and *then calculate* what the system's past evolution from S_i to S_f must have been, given the dynamical laws and the boundary conditions.)

What Frisch has in mind seems to be the following difference between the initial and the final state. Under the assumption that the system is closed between t_i and t_f we can prepare the initial state by intervening on it directly at t_i, but we can only prepare the final state by first preparing the initial state and then letting it evolve into the final state. If we were to prepare the final state directly by intervening on it at t_f then the system would not be closed anymore, which contradicts the assumption. This difference in how we can prepare the initial and final states provides us with a temporal asymmetry and therefore with a difference between past and future, according to Frisch.

Frisch then continues to argue that state preparation is a kind of intervention that falls under the interventionist account of causation by Woodward (2003):

> In particular, if S_f is an effect of S_i, then according to an interventionist account of causation there are two ways by which one can intervene on the system to set S_f to a particular value: first, we can intervene on S_i, which in turn will affect the value of S_f; or, second, we can intervene directly on S_f [...]. (Frisch 2010, p. 82)

To identify state preparations with causal interventions then enables Frisch to explain the asymmetry of state preparation with the asymmetry of causation. Finally, to complete his argument, by inference to the best explanation there must be causal relations in physics.

I do not agree with Frisch's argument and have a couple of reasons why. First, one might wonder whether causation really is the best explanation for the asymmetry of state preparation. According to Frisch (2010, p. 83) "there is no other fully worked out and equally as successful non-causal alternative explanation of the asymmetry." But what about a theory of time according to which time flows from past to future and only the present exists? This would do the same job by giving a reason why we cannot intervene on the final state directly without violating the closed system constraint. Granted, there might be other reasons for not believing into a flow of time, but there might just be the same reasons why many refuse to believe in time asymmetric causation. In any case, one would wish to have more reasons from Frisch to accept the explanatory relevance of causation. Second, Frisch does not make appropriate use of Woodward's interventionist theory. As can be seen from the quote in the previous paragraph, according to Frisch, if there is a causal relation between S_i and S_f, then an intervention on S_f can either go via S_i or be on S_f directly. However, according to definition IV.I3 in Woodward (2003, p. 98) a well defined intervention cannot act on S_f on a path that does not go through S_i (cf. section 3.3.4 above). This might seem like a nitpicking point, but it is crucial to Frisch's account that there is an asymmetry between certain ways of intervention that can be identified as causal interventions such that then the asymmetry can be identified as a causal asymmetry.

Most damaging to Frisch's account however is that upon closer examination the asymmetry of state preparation is nothing else than the temporal asymmetry that we can only intervene on the future but not the past. Which in turn makes Frisch vulnerable to Reichenbach's point that this statement is a mere tautology and therefore does not prove anything. Again, the asymmetry of state preparation according to Frisch consists in that we can intervene on S_i directly to prepare it, but not on S_f, given that the system is closed for times between S_i and S_f. The problem with that is that it seems to presuppose that time is directed from t_i to t_f. Because if this were not the case, one could turn Frisch's example around and look at it from the other temporal direction, in which case we could intervene on S_f directly, but not on S_i. Frisch does not give any reason why only one temporal direction should be allowed. Being charitable, at this point one could make use of the temporal asymmetry of the facts, for which I have argued in the previous section, in which case only one temporal direction is the case. Nevertheless, even then Frisch's account does not solve the problem of nomic bidirectionality. Even if there is an asymmetry in preparation in that we can prepare S_i directly, but S_f only by first preparing S_i and letting it evolve into S_f, the problem remains that S_i and S_f nomically depend on each other. Hence, Frisch' reasoning might solve the problem of temporal symmetry, but not the problem of bi-directional nomic dependence. I will therefore not follow Frisch, but propose a different account for the asymmetry of causation, based on a localised understanding of determinism and interventions, in the next section.

4.3.3 Determinism localised

The argument from nomic bidirectionality hinges on the determinism of physics. However, determinism is not a self-explanatory term and there are various ways of understanding it. In what follows, I will argue in favour of a local definition of determinism. Building up on that, in the next section I will show that causation is compatible with local determinism.

A standard text on determinism is Earman's 1986 'A Primer on Determinism'.

Here we find a first definition that I call global determinism:

Global determinism: Let $\mathcal{W}$ be the set of all possible worlds that satisfy the laws of physics. Then world $W \in \mathcal{W}$ is deterministic iff, for every W and all times t, if $W(t) = W'(t)$, then $W(t') = W'(t')$. (see Earman 1986, 13)

However, while this definition tells us when a world is deterministic, the argument from nomic bidirectionality establishes a conflict between causation and a physical system. Hence, what we actually need to know is what it means for a physical system to be deterministic. I wish to put forward the following definition:

Local determinism: Let T be a theory. Let $\mathcal{M}_T$ be the set of all dynamically possible models that satisfy T, and $M \in \mathcal{M}_T$. Let $M(t) \subset M$ be the model at a time t, and $s(t) \subseteq M(t)$ be the state of a system. Then M is deterministic iff, for every s and all times t, there exists within T a unique bijective map Φ_T such that $\Phi_T : s(t) \to s'(t')$.

What characterises a model is notoriously hard to define. However, for present purposes we might understand it simply as a set of states that solve the laws of T. Accordingly, I take the term 'physical system' used earlier to be synonymous with model. The term 'state' is ambiguous as it can refer either to an element of a model, say the state of a single particle, or a subset, say the state of multiple particles. The map Φ_T is determined by the dynamical equations of T, such that if a state interacts with other states of the model, then this interaction must be included in the dynamical equations.

Global determinism follows from local determinism, if s is a state of the whole world. Thus local determinism can recover the more common global determinism as a special case. Sometimes determinism is also defined as theory determinism: "We say that a theory is deterministic if and only if: any two of its models that agree at a time t on the state of their objects, also agree at all times future to t." (Butterfield 1998, 37) Again, local determinism recovers theory determinism, iff all $M \in \mathcal{M}_T$ are deterministic. However, as I will explain now, there are good reasons to prefer local determinism over the other formulations.

First, if the definition of determinism should be applicable to theories at all, then we might actually have no other choice than abandoning global determinism. Physical theories in general do not describe worlds, but systems. These systems can in principle be as large as what we call a world, but they do not have to be; for the theory this is irrelevant. Hence, the term 'world' is not well defined by a theory. Furthermore, global determinism is practically inapplicable. Given that we have no theory of everything and given our limited capacities, we cannot give a description of the whole universe, and whether this will ever be possible is an open question. Hence, global determinism is void in the context of today's theories. With Wüthrich (2011, 368) we might see this simply as a consequence of naturalism, that is, a focus on the interpretation of scientific theories, rather than possible worlds.

Second, if for the sake of the argument we apply global determinism to a theory, then presumably the theory would describe nomic dependencies between states of the whole universe. As a consequence, a valid formulation of the argument from nomic bidirectionality must assume that there are causal relations between states of the whole universe. Whether there are these global causal relations is a contentious matter, but what is clear is that most theories of causation are applicable to smaller objects, and intuitively most would agree that there are causal relations between smaller objects if there are causal relations at all. Or in other words, global causal relations, if they exist, are arguably not fundamental, but build up from local causal relations. Insisting on global determinism would rule out any approach to answer the question on the local level by definition. Local determinism, on the other hand, is more liberal, since a model can in principle be the whole universe, but not always is.

Third, physical theories such as Newtonian physics or quantum mechanics have models describing deterministic as well as models describing indeterministic processes (see Earman 2007; Norton 2008; Wüthrich 2011). As a consequence, one might call these theories indeterministic as a whole, however, a balanced approach seems more feasible. If a theory is deterministic or not relative to a model, then determinism of a theory should be defined relative to a class of models. This leaves no other choice than to abandon global determinism as well as theory determinism.

Contrary to that, Hoefer (2010, sec. 2.2) argues in favour of global determinism, for otherwise there could be influences from outside the local system that would lead to a spurious breakdown of determinism. Thus, it could be the case that $M(t) = M'(t)$, but $M(t') \neq M'(t')$, because there was an influence on M from outside the system described in the model. Of course, this is not what is meant by indeterminism. However, as Butterfield (1998, 36-37) explains, this problem is easily avoided if a model is regarded as representing an isolated system that has no interaction with other systems, as it is commonly done in physics. It might be objected that real systems are never completely closed. However, if being closed is not taken as a condition for the systems in question, then, following Hoefer, we usually would not have deterministic systems, which in turn would make the directionality argument invalid and thus brighten the prospects of squaring causation and physics rather than dampening them.

4.3.4 Reconciling causation and physics

As Eagle (2007, p. 158) proposes, the problem of bidirectional nomic dependence could easily be solved by relying on the temporal asymmetry of causal processes, for which I argued in section above 4.3.1. According to Eagle, it could be maintained that causation is a symmetric dependence relation and that the asymmetry of causation is provided by the temporal asymmetry. Given that physical processes in the world are temporally asymmetric, causation would then obviously be compatible with the bi-directional nomic dependence of physics, since causation itself would be a bidirectional dependence relation. Hence, causes would as much determine their effects, as effects would determine their causes. Only, because of the temporal asymmetry of processes in the world, one is contingently realised before the other in time, which makes it appear as if causes determine their effects, but not vice versa. Even though I believe this is a promising line of thought, I do not wish to follow it here. Historically it seems more accurate to understand causation itself as being asymmetric in two ways, temporally and causally (cf. my discussion in section 3.3.5), and following Eagle's idea is in danger of changing the concept, rather than illuminating it (cf. my discussion of Ney 2009 in section 4.2.3). Instead, in what follows I wish to argue that, presupposing the factual temporal asymmetry of physics, despite the nomic bidirectionality of physical laws, processes can show nomic directionality that is suitable for causation.

It is Eagle again, who pointed towards a possible reply to Russell's argument that he, however, did not work out. According to Eagle (2007, p. 158), a system being deterministic does not preclude the possibility of asymmetric determination relations:

> Even in a deterministic system, it is possible that both (i) all trajectories which feature this event type c as part of some global state s at t have some further type of event e as a feature of a state $s\,'$ at $t\,'$, and (ii) not every global state that features e at some time lies on a trajectory which involves some past state that features c. So the occurrence of c determines the occurrence of e in a way that e doesn't determine the occurrence of c. Focusing attention on events of a purely local character, rather than the entire state of the system, might very well give us an asymmetry of determination between particular events.

Thus, Eagle's idea is that in a system *S,* which is smaller than the whole universe, it is compatible with determinism that every state $S_1(t_1)$ evolves deterministically into a later state $S_2(t_2)$, but not every state $S_2(t_2)$ has evolved from an earlier state $S_1(t_1)$. Unfortunately, Eagle merely highlights the mere possibility of such a structure, and does not explicate further, how it could be realised. This will be my task in what follows.

Already Earman has presented a similar possibility and concluded that: "causal directionality is not incompatible with determinism" (Earman 1976, 18). Accordingly, there could be an intervention from God, or more realistically, an force from outside the system, that, like the Lorentz-Dirac equation changes the state of the system before it is turned on. This intervention, however, would typically violate the laws of T, and is in conflict with the condition applied here that the deterministic system is closed. Why is this a problem? The reasoning below will argue that for a given theory T determinism and causation can be reconciled. Changing that theory is in danger of being ad hoc. The closed system condition, on the other hand, is needed to prevent spurious breakdowns of determinism, and it seems reasonable to demand of a causal interpretation that it can include closed systems. Furthermore, Earman's argument relies on the existence of such a force, for which there currently is no empirical evidence.

As a first step in concretising Eagle's idea, I wish to make use of work on asymmetric causal structures, which has long been popular in economics, but gained less attention in philosophy. In particular Simon & Rescher (1966) have worked out a precise notion of causal structures that can serve as the blueprint to spell out Eagle's idea, and show how bidirectional nomic dependence can be compatible with causation. According to Simon & Rescher (1966, p. 324) a "causal relation is not a relation between values of variables, but a function of one variable (the cause) on to another (the effect)." The basis of a causal structure is a set of one or more functions that has to be self-contained, that is, it has to contain as many variables as functions. A larger self-contained structure can show a kind of asymmetry in that some functions can be solved without first solving others, but not vice versa. This will become clearer with an example, even though I can hardly pay justice to all the subtleties of Simon & Rescher's idea.[33]

Suppose that two billiard balls collide elastically at time t_0. Their momenta before the collision are fixed to $\bar{p}_1(t < t_0)$ and $\bar{p}_2(t < t_0)$. This determines the momenta after the collision, $p_1(t > t_0)$ and $p_2(t > t_0)$. The bar on the former two momenta indicates that they are exogenously fixed to certain values, while the latter two momenta are variables. By the conservation of momenta, we have $\bar{p}_1(t < t_0) + \bar{p}_2(t < t_0) = p_1(t > t_0) + p_2(t > t_0)$. Following the notation in Simon & Rescher, this set of four variables can be put into four functions, together forming a self-contained structure.

1. $f_1[\bar{p}_1(t < t_0)] = 0$
2. $f_2[\bar{p}_2(t < t_0)] = 0$
3. $f_3[\bar{p}_1(t < t_0), \bar{p}_2(t < t_0), p_1(t > t_0)] = 0$
4. $f_4[\bar{p}_1(t < t_0), \bar{p}_2(t < t_0), p_2(t > t_0)] = 0$

The first two equations signify that the values of $\bar{p}_1(t < t_0)$ and $\bar{p}_2(t < t_0)$ are pre-set to certain values, say, by the experimenter. The latter two equations signify that the values $p_1(t > t_0)$ and $p_2(t > t_0)$ can be calculated from the momenta before the collision. This system is directed in that equation f_1 and f_2 can be solved before solving the other two equations, simply by putting in the values that the momenta are fixed to before the collision. The momenta after the collision can only be calculated if we already know the momenta before the collisions. Thus, f_3 and f_4 can only be solved, if f_1 and f_2 are already solved. In the notation of Simon & Rescher, this directionality can be captured by two diagrams:

[33] I have chosen Simon & Rescher as the reference work here, because they address the topic of directionality most explicitly. A much more detailed account of defining causal relations in their spirit can for example be found in Pearl (2000). It might be noted that Pearl discusses Simon & Rescher's approach in chapter 7.2.5 and finds that "[t]he asymmetry that characterizes causal relationships in no way conflicts with the symmetry of physical equations." (Pearl 2000, p. 228) I wish to point out, that I am not trying to reduce causation to structural equations (see a critique of such an approach in Hall 2007 and a reply in Hitchcock 2009).

$$
\begin{array}{ccc}
f_1 & & \\
& \searrow & \begin{cases} f_3 \\ f_4 \end{cases} \\
& \nearrow & \\
f_2 & &
\end{array}
,\qquad
\begin{array}{ccc}
\bar{p}_1 & & \\
& \searrow & \begin{cases} p_1 \\ p_2 \end{cases} \\
& \nearrow & \\
\bar{p}_2 & &
\end{array}
$$

We can say that $\bar{p}_1$ and $\bar{p}_2$ are exogenous variables, while p_1 and p_2 are dependent variables. Given this directionality, Simon & Rescher feel confident enough to say the exogenous variables are causes, while the dependent variables are effects. Hence, in the above example one can conclude that the momenta $\bar{p}_1$ and $\bar{p}_2$ jointly cause p_1 and p_2.

To clarify where the advantages of Simon & Rescher's account lie, it is worth taking a quick look on another attempt to formalise causal relations. According to Frisch (2010, sec. 3), the causal relation can be represented by defining an asymmetric, transitive and non-circular relation C between two states S of a system. "*C is interpreted as the causal relation: $S(t_2)$ bears C to $S(t_1)$ exactly if $S(t_1)$ is a cause of $S(t_2)$. If two states do not stand in relation C then they are not causally related. The result is a class of what I want to call potential causal models of a theory.*" (Frisch 2010, p. 79) I wish to point out one problem with Frisch's account that shows the superiority of Simon & Rescher. As the latter observe, a function $f(x) = y$ usually possesses an inverse $f^{-1}(y) = x$. If now this function is in some way supposed to describe a causal relation, x causes y, then it seems plausible that the inverse function shows the inverse causal relation, y causes x. Now, while this is unproblematic for functions, the directionality of causation should prohibit the causal relation from having an inverse. This problem is recognised by Simon & Rescher and they solve it with the notion of self-contained systems; the asymmetric structure a self-contained system exhibits cannot be inverted in that way. The same, however, is not true for Frisch's causal relation C. Prima facie there is no reason why C should not have an inverse C^{-1} that is also a causal relation; even the asymmetry of C does not preclude this (which is just to repeat the point that asymmetry is not sufficient for directionality). There then seems to be no grounds to decide which relation, C or C^{-1}, is the real causal relation. Now, it might be possible for Frisch to add the further stipulation that C does not possess an inverse or that C, but not C^{-1} is a causal relation. However, in the present context this would not bring us any further, since this move would not provide any understanding about where this directionality comes from. Contrary to that, the directionality follows naturally from Simon & Rescher's account.

It is clear that simply stipulating a causal relation, as Frisch does, cannot serve as an argument for the existence of causal relations, and, to be fair, it is not meant as such by Frisch. Needless to say, the same is true for Simon & Rescher; that they chose to call their asymmetric relation causal does not mean that it actually is. Thus, a further argument has to be made to underpin Simon & Rescher's view that they are indeed describing causal relations. The point I wish to make is that this argument comes from Woodward's interventionism.

An obvious objection against the billiard example is that even though the way I described the collision forms a self-contained system, this is not sufficient to establish that the process is a causal one. The problem is that in this example the choice of $\bar{p}_1$ and $\bar{p}_2$ as exogenous and p_1 and p_2 as dependent variables is arbitrary. For what reason should the former have any priority over the latter? I believe this objection can be answered with the help of the interventionist theory by Woodward, which provides a structure that is that of a self-contained system with directional dependence relations. Recalling the discussion from section 3.3.4, it is part of the definition of an intervention that it has to be exogenous to a causal process, that is, if there is an alleged causal process in which X causes Y, then an intervention I has to come from outside of this process and break any previous influences on X, such that the variable X that causes Y is entirely set by I (cf. Woodward 2003, p. 94). Back to Simon & Rescher's formalism, this is what provides the priority of some variables over others; only if we already know the value of the intervention we can calculate the effect, but from the effect we cannot calculate whether there was an exogenous intervention or whether it has been caused solely by an endogenous cause. Following the notation of Simon & Rescher (1966, sec. 6), interventions, I_i, can be made explicit in the formalism, which, in the billiard case gives the following structure:

$$\begin{array}{lll} I_1 \to \bar{p}_1 & & \\ & \searrow & \\ & & \left\{\begin{array}{l} p_1 \\ p_2 \end{array}\right. \\ & \nearrow & \\ I_2 \to \bar{p}_2 & & \end{array}$$

What we have so far is the result that even a system of functional relations can be directed, and that this directionality can be understood as causal. It is clear, however, that this explication is not a reduction of **(CD)** to non-causal directionality. However, such a reduction was neither aimed at, nor will it be necessary for the overall argument. Prima facie, **(CD)** can be compared with physics even if it is a genuinely causal concept. Analogously, for example, one can ask whether there are laws in physics, even if the term 'law' is not fully reduced to non-nomological concepts. What is necessary, however, is some understanding of **(CD)** that we can work with, that is more than the statement 'causes are causes', and the above analysis provides such an understanding.

The next and final step now is to show that this directionality is compatible with determinism. To answer this question, I wish to build up on local determinism in order to formalise Eagle's idea as follows:

Asymmetric determination: Let T be a deterministic theory with models $\mathcal{M}_T$. Let $s_i(t_1) \subseteq M_i$ and $s_i(t_2) \subseteq M_i$, $t_1 < t_2$, be two states of the model $M_i \in \mathcal{M}_T$. Both states stand in an asymmetric determination relation iff, there exists a model $M_j \in \mathcal{M}_T$ with a state $s_j(t) \subseteq M_j$ such that $s_i(t) = s_j(t)$ for all $t \geq t_2$ and $s_i(t) \neq s_j(t)$ for all $t \leq t_1$.

Thus, there are two models that are isomorphic after t_2, but not before t_1. Why is this asymmetric determination? Because given any state $s(t \leq t_1)$, using the laws of T, we can find a unique bijective map such that $\Phi_T : s(t \leq t_1) \rightarrow s(t \geq t_2)$, but there is no unique bijective map for the converse, that is, given a state $s(t \geq t_2)$. The reason for this is that a state at $t \geq t_2$ can have evolved from either of the non-identical states $s_i(t \leq t_1)$ or $s_j(t \leq t_1)$. In conclusion, asymmetric determination is consistent with local determinism, given that each model M is deterministic and the asymmetry is a relation involving two different models. According to the earlier discussion, this amount to asymmetric nomic dependence, since from any $s_i(t \leq t_1)$ and $s_j(t \leq t_1)$ the later state $s(t \geq t_2)$ follows, but not vice versa.

The crucial step now is to ask in which sense these asymmetric dependence relations can be interpreted as asymmetric causal relations? Earlier, I have argued that causal directionality can be analysed into asymmetric dependence relations within complete structures. Furthermore, I explicated these structures in terms of an interventionist account of causation. The asymmetry we found consists in that the effect depends on exogenous intervention, but not the converse. With a local definition of determinism, we are now in a position to write down this asymmetry of interventions more formally, to see whether it is indeed consistent with local determinism.

Asymmetric causal structure: Let $I(t_1) \subset M_1$ be an intervention on a state $s_1(t_1)$ such that according to the laws of T at a later time t_2 we find the state $s_1(t_2) \subset M_1$, $t_1 < t_2$. Let $s_2(t_2) \subset M_2$ be a state that is isomorphic to $s_1(t_2)$, and $M_1(t > t_2)$ is isomorphic to $M_2(t > t_2)$, but $M_1(t < t_1)$ is not isomorphic to $M_2(t < t_1)$. If $M_1, M_2 \in M_T$, then T has models that establish an asymmetric causal structure.

In other words, the asymmetry consists in that both $M_1(t < t_1)$ and $M_2(t < t_1)$ determine the occurrence of $s_2(t_2)$, but the latter does not determine the occurrence of $I(t_1)$ or $s_2(t_1)$, given that both are compatible with the laws of T and dynamical evolutions involving $s_2(t_2)$. Or to say it in the terms of Simon & Rescher $s_2(t_2)$ is a function of $I(t_1)$ or $s_2(t_1)$, but not vice versa.

Perhaps the most obvious objection to the account here presented is that it works in both temporal directions and therefore fails to circle out one direction of nomic dependence over the other. Prima facie, the intervention can lie either in the past or the future of its effect, and there might even be symmetric cases with two interventions. So far, neither possibility can be excluded. However, at this point I wish to make use of the previously established result that processes in the world have a factual direction in time. This rules out interventions that lie in the future of their effect, since a past event cannot be manipulated by a future intervention when time has the corresponding direction. It would be question begging to rule out future to past interventions by merely appealing to the causal fact that we cannot intervene on the past. This is not what I propose. Rather, what is doing the work in ruling out future to past interventions is the temporal directionality of the physical facts, which was established independently of any causal concept.

Let me evaluate this point in a little more detail. If, for the sake of the argument, we assume that causal relations in both temporal directions are possible, then, according to Woodward's condition **(NC)**, the following two counterfactuals could be true: 1. 'If the intervention $I(t_1)$ was carried out, then the value of $X(t_2)$ would change.' 2. 'If the intervention $I(t_2)$ was carried out, then the value of $X(t_1)$ would change.' However, if contrary to that we assume that time is directed from t_1 to t_2, then whatever is done at t_2 will not change anything at t_1 and the second counterfactual would be false. This is not the result of an asymmetry of interventions, but an asymmetry of time.

Two clarifications: First, it is not the fact that one event comes before the other that makes the earlier the cause and the later the effect. Rather, it is the fact that interventions, that is, manipulating one variable by manipulating another, only work towards the future that establishes the causal order. Second, the causal directionality defended here is not the simple fact that 'we can only manipulate the future', but comes from an asymmetry between possible models. The temporal directionality of interventions is not used to establish the directionality of causation,[34] but merely to pick out one causal direction over the other.

Another objection might be that in the world only one possible model can be realised. Hence, while one might agree that there is an asymmetry between possible models, one might ask where this asymmetry is gone when only one model is actual. However, as Woodward (2003) has argued, there is a close link between interventions and counterfactuals. For a process to be causal, an intervention does not actually has to happen, it only has to be counterfactually possible.[35] Accordingly, for an asymmetric causal structure, it is not necessary that two models are realised. It suffices that an alternative model involving an intervention is possible according to the laws of T. What I must presuppose, however, is that all models that are considered have the same temporal direction.

I wish to finish this section by addressing a last worry. What we set out for was to analyse the asymmetry of causation and to ask whether this is compatible with physics. What we ended up with is an asymmetry between different models, different ways the world could have been. Isn't that too weak? Isn't that an asymmetry not of causation itself, contrary to the asymmetry of, say, bringing about? Ultimately I cannot dismiss these concerns, however, I wish to make a few comments. To the first question, causation is a vague concept, and different people can have different intuitions about it. It proved difficult to explicate the asymmetry of causation in a non-tautological way, and to provide an informative analysis naturally runs into the danger of not agreeing with someone's intuition. I can live with this sort of disagreement, because on the positive side we found at least one way to explicate the asymmetry that indeed is compatible with physics, which is better than incompatible tautologies. Whether this result is too weak may be decided by others. Concerning

[34] See the critique of such accounts in section 4.3.2.

[35] How these counterfactuals should be understood and can be evaluated has been precisely analysed by Woodward. However, see Psillos (2007) for a critique and the survey of objections in Menzies (2014, sec. 4.3).

the second question, causation is a concept comprising various different things, like cause, effect and a relation between them. Not everything that belongs to the concept is necessarily an intrinsic property of the objects or relations that belong to it. The temporal asymmetry of causation is actually a simple temporal relation between two events or objects that is incorporated into the concept 'causation', not a temporal asymmetry of causation itself. Analogously, the asymmetry between models can be understood as a part of the concept 'causation' without being intrinsic to any of its parts.

4.4 Conclusion

In this chapter I have argued against the widely held view that causation, as a directed relation, is incompatible with the symmetries found in physics. First of all, I have shown that time reversal symmetry of physics does not give rise to an argument against causation. The temporal asymmetry of causation means that causal processes have a direction. This is compatible with time reversal symmetry, since the latter, properly understood, only means that processes that are possible in one temporal direction are also possible in the other temporal direction; it does not mean that processes have no temporal direction at all. At the same time, it did not follow from here that processes actually have a temporal direction. Thus, I argued further that processes as observed in the world have properties sufficient to assign a temporal direction to them.

Furthermore, I have discussed a second argument that establishes a conflict between the determinism of physics and the unidirectional dependence of causation. Against this argument I have objected that interventionism together with the factual temporal direction of processes in the world is sufficient to establish unidirectional dependence relation that can further be formalised using the self-contained system approach.

Concordantly, the conclusion I wish to draw from this chapter is that physics is not as hostile towards causation as often believed. Time reversal symmetry and the apparent bidirectional nomic dependence of physics do not preclude the possibility to understand processes in physics as causal processes. The importance of this result cannot be underestimated, given that the arguments presented here are often the main reasons why philosophers of physics eschew causation and, if successful, would have rendered the present project much more difficult to carry out.

5 The quantum field theoretical description of interactions

5.1 Introduction

The central concepts of the CQT, as presented in chapter 2, are causal processes, explicated in terms of world lines of objects, and causal interactions, explicated as the intersection of world lines of causal processes. In the attempt of taking them into quantum physics, problems showed up corresponding to each one of the concepts. The notion of objects is unclear, because the problem of finding an ontology for QFT is unsolved. Furthermore, world lines become difficult to define, since the Heisenberg uncertainty relation forbids particles to have a sharp position and momentum at the same time and QFT seems to be a theory about fields, not point-like particles. Finally, intersections have to be regarded as quantum field theoretical interactions, which raises among others the problem of the existence of virtual particles.

However, while Dowe's CQT will continue to be a guiding idea in the background, following the methodology from section 3.2, I will not try to answer its problems directly. Rather, in this chapter I will analyse QFT in order to find some traits that look promising as a reduction basis for the platitudes from chapter 3. To begin with, I will take a more practice oriented look at QFT, especially at experiments, to assess whether a simple picture of particles on world lines that interact when they intersect can be upheld. The result will be negative, which in turn motivates a closer look at the theoretical description of interactions in QFT, which will be done in the remainder of this chapter.

I start by presenting three case studies of how interactions are described in theoretical physics, in order to gain a better understanding for QFT. The first case study will be the semi-classical scenario of a quantum particle in a classical field. The second one will be from non-relativistic quantum physics, and the third will be from quantum electrodynamics. By proceeding stepwise from classical to relativistic quantum physics, it will hopefully be possible to understand what is general throughout about how interactions are described, but also what is special to QFT.

After the case studies, I will turn to particular aspects of the theoretical description of interactions in QFT that show up at different points in the case studies, and discuss them in six separate sections. Even though all of these aspects come together in QFT, among themselves they are mostly unrelated. The six sections thus cannot build up on each other, and it might seem as if their topics were chosen randomly. However, once I return to the issue of causation in the later chapters, it will hopefully become clear why their topics were chosen and what binds them together.

During this chapter, it will surface more and more that, contrary to Dowe, a promising reduction basis for QFT lies not in controversial interpretations about particles and intersections, but in general principles of QFT. In particular group structure, locality and local conservation will crystallise as the central concepts that a causal interpretation might be based on. However, the concrete proposal for a new theory of causation that refers back to these principles has to wait until chapter 7.

5.2 World lines

5.2.1 Particles and fields in classical mechanics

Intuitively, a very accessible picture of how causal processes in physics might look like is that of small objects travelling through space and eventually meeting and interacting at certain points. Consequently, if there are such things as, e.g., small point particles, they may serve as the perfect reduction basis for the causal platitudes. Also, this seems to be the idea for which the CQT is coined, namely for the mechanics of localised particles. Dowe's definition of the CQT relies on objects, world lines and intersections of world lines, which can most easily be conceived of as tracking point-like particles or objects like billiard balls that eventually bump into one another and thereby transfer some energy-momentum or other conserved quantities. However, in QFT this picture can hardly be upheld.

At least on the surface, QFT is a field theory, where a field is a function that assigns to each spacetime point a certain value, and as a consequence of Malament's (1996) theorem it is very clear that it is incompatible with an ontology of exactly localised particles. Even though Dowe realises that sometimes causal influences are transmitted by force fields, this does in no way alter his theory, for he holds that "fields should be construed as objects for the purpose of this theory." (Dowe 2000, p. 53) Consequently, fields do not pose a special case that should be treated differently from any other objects. However, it is not obvious that things are that easy. For example fields can be extended over all of spacetime, have very different values for their field strength at different points in spacetime, and the status of, e.g., energy is not always easy to define in field theories. In trying to clarify how a theory of causation might handle fields, I will proceed stepwise and first treat classical field theories before looking at the quantum case in the next section.

From what has just been said, one can expect that Dowe's CQT fits perfectly with Newtonian mechanics. The latter describes point particles that can be fully characterised by their position in an 6N-dimensional phase space. Forgetting about action at a distance for the moment, it is obvious what world lines are and where they intersect. However, Newtonian mechanics need not be regarded as describing point particles. Indeed Stein (1970, p. 264) says that "I do not think there is a clearly delimited fundamental kind, or form, or structure, which distinguishes field from non-field theories." Stein finds this view already implicit in Newton's theory and explicates it with the example of a two-valued field, an idea which later Redhead

(1982) made more precise. The thrust is that in a particle theory the spatial variable, which gives the location of the particle, is replaced by a field which assigns to each space point one of two values: either 'yes' or 'no', either there is a particle at a point or not. Accordingly, the dynamics of a particle can be described by tracing the field value 'yes' over time. Of course, this method is not only possible for point particles, but as well for extended objects. The field will then have the value 'yes' not only at a point, but in a certain region.

Given the general possibility of a translation of particle into field theories, it is plausible to assume that also the CQT can be translated into field theoretical terminology; not in Dowe's crude way of taking the whole field as a single object, but more finely grained. This step is rather trivial in that Dowe's objects are now the part of the field that have a positive value and a world line is the path in spacetime that the positive part of the field takes. Furthermore, intersections of world lines happen, when two fields have a positive value at the same spacetime point. This framework also treats particles and, e.g., the electromagnetic field essentially alike. Field theories are therefore prima facie unproblematic for transference theories of causation

However, the situation is not that simple when one changes from classical to quantum fields. In QFT it is rather unclear what the field values depict and whether there are world lines that do not extend over the whole field. Having said that, the result of the next section will not be much different from Stein's and Redhead's way to regard classical mechanics as a field theory. Nevertheless, as will become apparent later, world lines are not the best reduction basis for causation, and I will argue for a different approach.

5.2.2 Local quantum fields

For the attempt to take over the reasoning from the last section into QFT, immediately several problems arise. First, fields in QFT are typically defined over all of spacetime and have a value different from zero everywhere. In practice often cutoffs are introduced, viz., fields are only evaluated in a certain spacetime region, and it is assumed that the fields fall off quickly towards infinity. Nevertheless, if again the non-zero part of the field is regarded as constituting the world line, it follows that world lines will typically extend over all of spacetime. Even worse, if there is more than one field, then the corresponding world lines will constantly intersect and it is not clear how the unintuitive conclusion could be avoided that everything interacts causally with everything at all places and times.

The second problem is that it is not even clear whether it makes sense to interpret quantum field values as world lines. In the field formulation of Newtonian mechanics from the last section, a positive field value can be interpreted as the field being 'impenetrable' at that point, which means that there is a particle (cf. Stein 1970, p. 277). Force fields in classical physics assign a field strength to every spacetime point, which gives the force that a test particle would feel at that point. To the contrary, fields in QFT only assign operators to every spacetime point, for which an equally direct interpretation seems to be impossible. This situation is made worse by the fact that there is no position operator in QFT, which could be interpreted as giving the probability for there being an object at point x in space (cf. Ticciati 1999, ch. 1.6). This makes Dieks (2001) doubt whether the manifold, on which the quantum field is defined can rightly be called spacetime at all. Another reaction is that of Teller (1995) who tries to give operators at each spacetime point a direct physical interpretation as determinable properties. However, it can be doubted whether Teller's approach is successful (cf. Wayne 2002; Fleming 2002). As a result, even if the first problem can be answered and world lines be defined with the help of the field, it still remains unclear whether the world lines can be regarded as belonging to physical objects or not.

The most immediate, but naive, answer to the first problem might be to exploit whether fields can be defined in such a way that they only have a non-zero value within a spacetime region $\mathcal{O}$ and zero in the complement region $\mathcal{O}'$, that is the so-called Newton-Wigner localisation. This solution however fails, since it is in conflict with special relativity, as Malament (1996) shows.[1] This problem is evaded by a proposal due to Knight (1961).[2] He first points out that all measurable quantities in QFT are the results of products of fields acting on a certain state:

$$\langle a|\hat{\phi}(x_1), \ldots, \hat{\phi}(x_n)|a\rangle \ .$$

Based thereon, Knight defines the state $|a\rangle$ of a field $\hat{\phi}(x)$ as strictly localised in a region $\mathcal{O}$, if outside of that region, in $\mathcal{O}'$, it is the vacuum state $|0\rangle$, that is:

$$\langle a|\hat{\phi}(x_i)|a\rangle = \langle 0|\hat{\phi}(x_i)|0\rangle \quad \text{for all } x_i \in \mathcal{O}' \ .$$

Here a slight ambiguity arises as to whether we say that either the state or the field is localised. But nothing depends on it, since from the definition it is clear what localisation means and mathematically it can be constructed either by a suitable manipulation of the state or the field.

However, as Haag & Swieca (1965) point out, the definition of Knight alone is not tenable, because of the Reeh-Schlieder theorem. The latter exploits the entanglement of the vacuum state and has as a consequence that "we can approximate any state vector by linear combinations of vectors describing states 'strictly localized' in some region $\mathcal{O}$." (Haag 1996, p. 254) Therefore, states that satisfy Knight's requirement will in general not be localised in $\mathcal{O}$. On the other hand, the correlations of the vacuum state fall off very quickly with distance and to use them to create states over long distances requires "prohibitively high-energy states" (Wallace 2001, p. 10)

[1] Fleming & Butterfield (1999) have responded to Malament that Newton-Wigner localisation is possible, if localisation is always defined with respect to a spacelike hyperplane. This approach, however, has been criticised by Halvorson (2001).

[2] It should be noted that the definitions of Knight (1961) and, later to be discussed, Haag (1996) and Wallace (2001) are strictly only shown to be valid for massive bosonic scalar fields, but it can be expected that a generalisation to other fields is in principle unproblematic (cf. Knight 1961, p. 460 and Wallace 2001, p. 41).

This motivates that we weaken strict localisation to what Haag & Swieca (1965) call 'essential localisation' and Wallace (2001) calls 'effective localisation'. The idea is not to require that a state is the vacuum state outside of its region of localisation, but only that it approximates the vacuum state. If the probability for a state to be localised in a sphere around region $\mathcal{O}$ falls off sufficiently fast, one can choose a certain distance from $\mathcal{O}$ after which the probability is extremely small. Wallace (2001, p. 10) puts these considerations together to the following definition:

> Effective localisation (qualitative form): A state $|a\rangle$ is effectively localised in a spatial region Σ_i iff for any function $\hat{f}$ of field operators $\hat{\phi}$, $\hat{\pi}$, $\langle a|\hat{f}|a\rangle - \langle 0|\hat{f}|0\rangle$ is negligibly small when $\hat{f}$ is evaluated for field operators outside Σ_i, compared to its values when evaluated for field operators within Σ_i. [notation changed]

Now, the definition after which point exactly the probability for localisation outside of Σ_i is negligible is to some extent arbitrary, and I will say more on the justification for this form of localisation in the next section. For the moment it is sufficient that, following Wallace (2001, p. 27 f.), from the rate that the entanglement of the vacuum state decreases, we can expect the diameter of Σ_i to be on a length scale of the order of the Compton wavelength. As also stressed by Wallace (2001, p. 30), an important property of effectively localised states is that they remain at least approximately localised under time translations.

Coming now back to the initial problem of how the notion of world lines might be explicated in QFT, it is fairly obvious how effective localisation helps: A world line is the spacetime region $\mathcal{O}$ in which a field $\hat{\phi}$ is effectively localised. However, one might object that this definition of world lines is not very enlightening, since operator valued fields are no physical entities and therefore cannot have a world line in spacetime. This is the second problem, stated at the beginning.

This concern is certainly well justified, but that a quantum field assigns operators to each spacetime point is not all what one can say about that field. In general, "[a]ll quantum field theories [...] model localization by making observables dependent on position in spacetime" (Halvorson & Clifton 2002, 18). Therefore any observable $\hat{O}(x)$, defined at spacetime point x, with a non-vanishing expectation value is sufficient to refer to an entity localised at x.[3] Another virtue of the quantum field is that it specifies the energy-momentum of a state. This usually is calculated by letting the Hamilton operator, which is a product of creation and annihilation operators, act on the vacuum state. For an effectively localised state it trivially follows that (nearly) the whole energy-momentum is contained in the region where the state is localised. Now this will not ease someone who is concerned about ontology, like, e.g., Teller, because it does not tell us what it is that is localised. Nevertheless, since my main focus lies on causation, knowing that something is localised, whatever it is, is a major step forward.

In conclusion, even in QFT it is possible to define localised states with a localised energy-momentum. Furthermore, given that the CQT describes causation with the

[3] More precisely, operators in QFT are smeared out in spacetime.

help of world lines and the exchange of energy-momentum, the fact that we can define world lines that have an associated energy-momentum in QFT could serve as the starting point for some kind of transference theory in the context of QFT. Having said that, this discussion only treated free fields, and the remaining task is now to take a closer look on interactions, but before doing that, another serious worry needs to be discussed.

I take it that Wallace has chosen the adjective 'effective' for his version of localisation with care, since it has some pragmatist connotations. Indeed, Wallace's main concern is to explain the emergence of small particles through localised states. He does not claim that what there is are particles. At least as presented here, effective localisation has an ad hoc character and one might well question what justifies the claim that a state is localised in a region, even though the probability to find the state outside that region, though very small, is non-zero. As already mentioned, the choice of the region's size in which the state is said to be localised is prima facie arbitrary. A possible answer to this question is what I will engage with in the next section.

5.2.3 Lessons from experiments

The sole possibility to theoretically construct effective localised states is not enough to take them seriously. What ultimately grounds localised states must be physical practice, that is, experiments and their theoretical description. Effectively localised states have to be shown as actually occurring and not only as a possibility. Even though it will not suffice to only point towards measurement outcomes, since effective localisation is an assumption about states and not phenomena, a clear picture of how experimental phenomena come about will help to make inferences to the states that come before them. So what can we learn from experiments?

The primary use of QFT is the description of scattering events. Scattering experiments start with the production of particles.[4] Since these particles are supposed to be accelerated by magnets, they have to be charged. Electrons and ions are the particles that generally are produced first. There are a large number of different kinds of devices that can be used to produce them, but the fundamental processes are always more or less the same. Electrons are usually emitted from surfaces by either heating up the surface, putting a high voltage on it or bombarding it with photons. Protons on the other hand are extracted out of a plasma, e.g., by bombarding it with high energy electrons. All of these methods have in common that they can be designed such that the energy of the produced particles can be controlled extremely well (cf. Hill 1994). Any other particles can subsequently be produced from collisions of electrons or ions (cf. Grupen & Schwartz 2008, p. 85). These particles are then accelerated in a magnetic guiding field to very high energies, before being directed on a target.

[4] In the present context the term 'particle' is merely a practical placeholder for whatever actually exists, localised or not, and should not be understood literally.

After the collision the resulting particles are led into detectors.[5] The evolution of detectors started with scintillation materials and counting light flashes with the naked eye; but has now reached enormous dimensions with detectors such as the ATLAS at CERN, where signals are not anymore read out optically, but only electronically. Even though the size and technical complexity of detectors changed, the (sub-)atomic processes that are used in particle detection have not. Detection of particles is only possible if the particles interact with the detector in a way that changes the state of the detector and can be recorded, that is, some amount of energy has to go from the particle to the detector (cf. Maxwell 1988, p. 20). In principle, each of the four known forces of nature is capable in many different reactions of serving this role (cf. Grupen & Schwartz 2008, p. 1). However, in practice mostly only two processes are used, namely, ionisation and emission of radiation. Both processes are based on the electromagnetic force, and therefore, neutral particles have to be converted through scattering within the detector into charged particles before they can be measured. Generally, measurements can be divided into non-destructive and destructive measurements. The former do not change the position and energy-momentum of the measured particle considerably, so that the particle can be used for further measurements, while in the latter the particle is destroyed and its energy is more or less completely absorbed in the detector or transferred to other particles.

Non-destructive methods are typically used to measure time, velocity and momentum. Time can be measured, e.g., by letting a charged particle excite atoms in a scintillation material, which then emits light while de-exciting. The emitted light can be led into a photomultiplier where it is transformed into an electric signal by using the photoelectric effect. Velocity can be measured in time-of-flight measurements by doing two time measurements on one particle at different places. Finally, momentum can be measured, e.g., in drift chambers set into an electric field. A charged particle will ionise atoms of the gas that lie on its track within the chamber. By measuring the ionisation current the track of the particle can be monitored and from the curvature of the track, due to the external electric field, the momentum can be calculated using the Lorentz force law.

On the contrary, the total energy of a particle can only be measured destructively. By using a solid, such as silicon, instead of gas, as detector material, charged particles will very quickly lose all their energies due to ionisation at low, and due to bremsstrahlung at high energies. While for low energy photons the photoelectric effect dominates, for high energy photons, electron-positron pair creation dominates. Strongly interacting hadrons on the other hand first have to be transformed through scattering into a shower of other particles of which a fraction that is proportional to the total energy is charged and can be measured by the above-mentioned methods.

Additionally, every detector measures the location of a particle just because the detector has a finite size. By separating detectors into small parts, e.g., small strip-

[5] Standard textbooks about measurement techniques in particle physics are, e.g., Grupen & Schwartz (2008) and Green (2000). Most of my presentation is derived from them, but see also Falkenburg (2007, ch. 3, 4) for a concise presentation.

like electrodes, very accurate location measurements can be made. Furthermore, types of particles can be identified by combining several different measurement methods. The ATLAS detector for example consists of eleven different detectors, one after the other (cf. LHC-Experiments-Committee 1999).

The theoretical means to describe measurement processes and outcomes used in textbooks and by experimental physicists are usually either taken from classical physics, like the Lorentz force law, or empirically justified approximations or simplifications of the fundamental quantum field theoretical description. As Falkenburg (2007, p. 96) notes: “Today, the theory of position measurement is interlocked with many dynamic laws which form a complicated piecemeal assemblage of classical and quantum theoretical assumptions.” For example the Bethe-Bloch formula, which is commonly used to describe the energy loss due to ionisation, or the Bethe-Heitler formula for bremsstrahlung do not work with operators and states in Fock space, as one might expect in QFT, but merely with more or less classical notions (even though it has to be kept in mind that properties like energy are not continuous, but quantised). Another example is Cerenkov radiation that is commonly described by using the classical Maxwell equations.

Both the Bethe-Bloch and the Bethe-Heitler formula, are based on the Born approximation for scattering events. Since the latter is only valid if the interaction is small, it is clear that what really goes on can only be described by QFT. The total ionisation energy loss along the track of a charged particle, which can be calculated at once with the Bethe-Bloch formula, is actually a series of distinct interactions of the particle with bound electrons on its way. For an exact quantum field theoretical calculation, each of these interactions has to be taken as a scattering process between the incoming charged particle and a bound electron. Radiation due to bremsstrahlung or deexcitation, for which the Maxwell equations can give useful results, have to be described as the interaction between a charged particle and the Coulomb field of an atom by which photons are created. For all of these processes a certain transition probability and a cross section can be calculated, using the methods of QFT.[6]

Even though part of classical physics are still in use to analyse the data, captured in detectors, the initial scattering event and its outcome, that is, the cross section can only be described and predicted by QFT. I will go into the details of interactions in section 5.3, at this point it is only important to emphasise the connection between localised states and localised measurement outcomes. The most important results of the previous discussion are that measurement outcomes are always fairly well localised events, which can be evaluated partly with notions from the mechanics of classical point particles. However, this by itself does not mean that the states that are measured have to be localised, because measurement outcomes, for example tracks in cloud chambers, are not properties of the state, but “properties of the interaction process” (Haag 1996, p. 309) of the state with the detector. The link to

[6] For a quantum field theoretical treatment of interactions used in measurements see Nagashima (2010), and for radiation from scattering events see Peskin & Schroeder (1995, ch. 6).

QFT is made by the fact that localised states are sufficient for localised measurement outcomes (cf. Haag 1996, p. 301; Wallace 2001, p. 12). The reason is that one property of interactions in QFT is that they happen locally (I will treat locality in sect. 5.3.4). The fact that measurement outcomes are localised can therefore be explained by the assumption that the states are localised.

This locality is also in accordance with the theoretical description of scattering events. Here the initial and final states are constructed as wave packets, that is, a superposition of plane waves with well defined momentum. In practice often simpler plane waves are used without culminating them into packets, but then it has to be assured by boundary conditions that initial states behave as wave packets, so that they do not interfere before and after the scattering. The result is then the same as when working with packets (cf. Knight 1961, p. 459; Peskin & Schroeder 1995, p. 102; Greiner & Reinhardt 2009, p. 2). The relation between plane waves, wave packets and effectively localised states is summarised by Knight (1961, p. 462) when he writes:

> In the theoretical treatment of the scattering of fields, the initial and final states are usually idealized to one-particle states of definite momentum. Such states are not localized, and cannot be produced by apparatus confined to a bounded region of space-time. The localization associated with production and detection is accounted for in single-particle scattering theory by using a wave-packet description of initial and final states. It is in this sense that the localized states defined here provide a certain field theoretic analog of the wave-packet description.

In conclusion, the reliance on effectively localised states in order to define world lines in QFT is justified on the one hand by the use of wave packets in the theoretical description of interactions, and on the other hand by the fact that measurement outcomes always show well localised events. Both these aspects are furthermore connected by the principle of locality.

However, it is important to stress that even though the above justifies us in thinking of entities in QFT as localised in certain conditions, there is certainly no constraint for them from QFT to always be localised in that way. It will therefore be necessary to have a closer look into the theoretical description of interaction that QFT provides, in order to see whether there is something more suitable and convincing than local world lines. This I will be doing in the next sections.

5.3 Interaction and intersection

In the last section I have argued that there might be a way to understand processes in QFT in terms of fairly well localised world lines. However, this picture relies on some special conditions that are not always given. Furthermore, it is still unclear how interactions between the entities that make up the world lines take place. Is there something like a 'meeting of world lines' in QFT that could serve as a basis to retain Dowe's CQT? In order to answer these problems and to obtain a clearer picture of processes in QFT it cannot be avoided to dig deeper into the technical details of that theory. It what follows, I will therefore first go through three case studies, which show how interactions are computed in physics, and then highlight in particular features of the last case study, on which I will rely later when returning to causation. The case studies build up on each other in so far as the first describes the coupling between a particle and a force field, the second the scattering of particles and the third will be the full relativistic treatment of particle scattering involving a force transmitting the influence. The treatment, however, will only make explicit what is heuristically useful for the overall argumentation towards physical causation and necessarily omit many steps and details.

5.3.1 The mathematical description of interactions

5.3.1.1 Case study 1: A charged (quantum) particle in a (classical) inhomogeneous electromagnetic field

In order to understand interaction in QFT it will be helpful to not start with QFT directly, but with non-relativistic quantum mechanics instead. This facilitates the task of introducing certain characteristics of the formalism, which also play a role in the full relativistic treatment of interactions. This first case study is a helium atom in an inhomogeneous electromagnetic field and is in particular introduced here to show that the interaction between a particle, that is the bound electron, and an external force field is reached by coupling both in the Hamiltonian (cf. Bransden & Joachain 2000, ch. 11). Also, a first glimpse on the role of group theory will be given.

The task of computing the interaction can be simplified greatly, if one is not interested in a full description of the process, but merely a good approximation of how the bound electron behaves. The first approximation is that in general electromagnetic fields contain a high number of photons and can therefore be treated as continuous and described by the classical Maxwell equations. Additionally, a single helium atom will typically only emit or absorb single photons at a time, which makes the action of the atom on the electromagnetic field negligibly small. Third, the fact that the electron of the helium atom is bound by the strong central potential of the nucleus turns the external electromagnetic field into a relatively small perturbation of the bound states of the electron and thus the process can be computed by using perturbation theory. Finally, the fact that the mass of the helium nucleus is much larger than that of the electron means that the action of the external field on the nucleus can be omitted as well, without losing too much precision.[7]

The classical electromagnetic field is described by the electric field strength $\vec{\mathcal{E}}$ and the magnetic field strength $\vec{\mathcal{B}}$, which behave according to Maxwell's equations, and

[7]For this reason, a free electron in an inhomogeneous electromagnetic field is actually more difficult to describe. However, what one observes in this case is that the wave function of the electron behaves to some extent akin to how a classical particle behaves when acted on by the Lorentz force (cf. Kennard 1931).

are generated by the scalar potential ϕ and the vector potential $\vec{A}$ by the following relations:

$$\vec{\mathcal{E}}(\vec{r},t) = -\vec{\nabla}\phi(\vec{r},t) - \frac{\partial}{\partial t}\vec{A}(\vec{r},t)$$

$$\vec{\mathcal{B}}(\vec{r},t) = \vec{\nabla} \times \vec{A}(\vec{r},t)\,.$$

$\vec{\mathcal{E}}$ and $\vec{\mathcal{B}}$ are invariant under the transformations $\vec{A} \to \vec{A} + \vec{\nabla}\chi(\vec{r},t)$ and $\phi \to \phi - {}^{\partial\chi(\vec{r},t)}/_{\partial t}$, where χ is any real differentiable function. As a consequence of this gauge invariance, one is free to work with the so-called Coulomb gauge, in which $\vec{\nabla} \cdot \vec{A} = 0$ and $\phi = 0$.

The vector potential is then described by the wave equation

$$\Delta\vec{A} - \frac{1}{c^2}\frac{\partial^2\vec{A}}{\partial t^2} = 0 \qquad (5.3.1)$$

which is solved by a plane wave of the form

$$\vec{A}(\vec{r},t) = A_0(\omega)\hat{\epsilon}\cos(\vec{k}\cdot\vec{r} - \omega t + \delta_\omega) \qquad (5.3.2)$$

where $|A_0(\omega)|$ is the amplitude with direction of the unit vector $\hat{\epsilon}$, ω the angular frequency, $\vec{k}$ the propagation vector and δ_ω a phase.

In the Coulomb gauge, $\vec{\mathcal{E}}$ and $\vec{\mathcal{B}}$ now reduce to

$$\vec{\mathcal{E}}(\vec{r},t) = -\omega A_0(\omega)\hat{\epsilon}\sin(\vec{k}\cdot\vec{r} - \omega t + \delta_\omega)$$

$$\vec{\mathcal{B}}(\vec{r},t) = A_0(\omega)(\vec{k}\times\hat{\epsilon})\sin(\vec{k}\cdot\vec{r} - \omega t + \delta_\omega)\,.$$

The energy density of the electromagnetic field is given by the expression

$$\frac{1}{2}(\epsilon_0\,|\mathcal{E}|^2 + |\mathcal{B}|^2/\mu_0) = \epsilon_0\mathcal{E}_0^2(\omega)\sin^2(\vec{k}\cdot\vec{r} - \omega t + \delta_\omega)\,.$$

Furthermore, taking $N(\omega)$ to be the number of photons with angular frequency ω in volume V, one can define the intensity of the electromagnetic field by

$$I(\omega) = \frac{\hbar\omega N(\omega)c}{V}\,. \qquad (5.3.3)$$

Now that the description of the electromagnetic field is clear, the Hamiltonian, which is the total energy, of the bound electron has to be found. Here one can start with the classical energy of a free particle of mass m, which is given as $E = {}^{\vec{p}^2}/_{2m}$. Adding the potential energy $-{}^{Ze^2}/_{(4\pi\epsilon_0 r)}$, resulting from the Coulomb potential of the helium nucleus, yields the Hamiltonian for the bound electron without an external field

$$H_0 = \frac{\vec{p}^2}{2\mu} - \frac{Ze^2}{4\pi\epsilon_0 r}$$

where $\mu = {}^{mM}/_{(m+M)}$ is the reduced mass coming from the electron mass m and the nucleus mass M.

The coupling of a particle to an external electromagnetic field on the other hand can be described by taking again the classical energy $E = {}^{\vec{p}^2}/_{2m}$ and making the substitution $E \to E - q\phi$, $\vec{p} \to \vec{p} - q\vec{A}$, where q is the charge, viz., by adding the potential energy of the electromagnetic field to the Hamiltonian of the particle. The Hamiltonian then reads as

$$H_{ex} = \frac{1}{2}\left(\vec{p} - q\vec{A}\right)^2 + q\phi\,.$$

The transition from classical physics to a semi-classical theory, in which the particle is described by quantum mechanics and the external field still by classical electrodynamics, is now made by promoting the energy and momentum to operators $E \to \hat{E} = i\hbar{}^{\partial}/_{\partial t}$ and $\vec{p} \to \hat{p} = -i\hbar\vec{\nabla}$. The total Hamiltonian operator $\hat{H} = \hat{H}_0 + \hat{H}_{ex}$ then is given by

$$\begin{aligned}\hat{H} =& \hat{H}_0 + \frac{1}{2}\left(\hat{p} - q\vec{A}\right)^2 + q\phi \\ =& -\frac{\hbar^2}{2\mu}\Delta - \frac{Ze^2}{4\pi\epsilon_0 r} - \frac{i\hbar e}{\mu}\vec{A}\cdot\vec{\nabla} + \frac{e^2}{2\mu}\vec{A}^2\end{aligned}$$

which in the last step contains already the simplification due to the Coulomb gauge. Consequently, the Schrödinger equation for the helium atom in an electromagnetic field is

$$i\hbar\frac{\partial}{\partial t}\Psi(\vec{r},t) = \hat{H}\Psi(\vec{r},t)\,. \tag{5.3.4}$$

As a side note, it is worth mentioning that the gauge transformations of the scalar and vector potential ϕ and $\vec{A}$ is equivalent to a multiplication of the wave function $\Psi(\vec{r},t)$ with a local time-dependent phase factor $e^{iq\chi(\vec{r},t)/\hbar}$. This can be seen by inserting the gauge transformations $\vec{A} \to \vec{A} + \vec{\nabla}\chi(\vec{r},t)$ and $\phi \to \phi$-${}^{\partial\chi(\vec{r},t)}/_{\partial t}$ into the Schrödinger equation 5.3.4, which gives

$$i\hbar\frac{\partial}{\partial t}\Psi(\vec{r},t) = \hat{H}_0 + \frac{1}{2}\left(\hat{p} - q\vec{A} - q\vec{\nabla}\chi(\vec{r},t)\right)^2 + q\phi - q\frac{\partial\chi(\vec{r},t)}{\partial t} \tag{5.3.5}$$

On the other hand, by computing

$$\begin{aligned}i\hbar\frac{\partial}{\partial t}e^{iq\chi(\vec{r},t)/\hbar}\Psi(\vec{r},t) =& e^{iq\chi(\vec{r},t)/\hbar}\left(i\hbar\frac{\partial}{\partial t} - q\frac{\partial}{\partial t}\chi(\vec{r},t)\right)\Psi(\vec{r},t) \\ \hat{p}\frac{\partial}{\partial t}e^{iq\chi(\vec{r},t)/\hbar}\Psi(\vec{r},t) =& e^{iq\chi(\vec{r},t)/\hbar}\left(\hat{p} + q\vec{\nabla}\chi(\vec{r},t)\right)\Psi(\vec{r},t)\end{aligned}$$

one can see that 5.3.5 is identical to

$$i\hbar\frac{\partial}{\partial t}e^{iq\chi(\vec{r},t)/\hbar}\Psi(\vec{r},t) = \hat{H}e^{iq\chi(\vec{r},t)/\hbar}\Psi(\vec{r},t)\,. \tag{5.3.6}$$

Here group theory enters explicitly, since the phase factor is a unitary scalar quantity and an electron a representation of the group U(1).[8]

Coming back now to the interaction, in the following we will work in the so-called Dirac or interaction picture, in which both the Hamiltonian and the wave function depend on time. The main problem here is that a time dependent Hamiltonian, as in 5.3.4 can in general not be solved exactly, but only via perturbation theory. However, to simplify this problem, two additional approximations will be used. The first one is the omission of the term $\vec{A}^2$ in the Hamiltonian. This is possible, since the perturbation, that is the external field, is treated as relatively weak, which makes the quadratic term small compared to the linear term. The second one is the dipole approximation, which uses the fact that the wavelength of the electromagnetic field is typically much larger than the size of the helium atom. As a consequence, the field can be treated as uniform over the atom and 5.3.2 reduces to

$$\vec{A}(t) = A_0(\omega)\hat{\epsilon}\cos\left(\omega t - \delta_\omega\right)\,.$$

Now, perturbation theory starts from the assumption that the total Hamiltonian of the system can be separated into a stationary and a time-dependent part

$$\hat{H} = \hat{H}_0 + \lambda\hat{H}_{ex}(t)$$

where λ will identify the different contributions to the perturbation series.

As a starting point, it is assumed that the eigenfunctions and eigenstates of the unperturbed Hamiltonian are known

$$\hat{H}_0\psi_k^{(0)} = E_k^{(0)}\psi_k^{(0)}$$

and the $\psi_k^{(0)}$ form a complete orthonormal set of eigenfunctions. This means that a general solution to the unperturbed Schrödinger equation can be expressed by a superposition these eigenfunctions

$$\Psi_0 = \sum_k c_k^{(0)}\psi_k^{(0)}\exp\left(-iE_k^{(0)}t/\hbar\right)$$

such that $\left|c_k^{(0)}\right|^2$ gives the probability for the system to be in eigenstate ψ_k^0.

Given that the total Hamiltonian is time-dependent, the energy of the system is not conserved and consequently the task is not to find corrections to the eigenvalues

[8] The argument is often taken to go in the opposite direction, such that the requirement of invariance under local phase transformations requires the introduction of gauge invariant potentials (cf. Schulten 2000, ch. 8). I will not discuss this question further, because the interpretation of group structure, which I present in section 5.3.2, does not hinge on the question why QFT is symmetric under certain transformation, but merely takes the symmetries as given.

$E_k^{(0)}$. Rather, one seeks an approximation of the perturbed wave function Ψ from the known $\psi_k^{(0)}$ (cf. Bransden & Joachain 2000, p. 432). One can then go on an calculate the probability that the previously unperturbed state $\Psi(t \leq 0)$ is found in state ψ_b after the perturbation has been switched on at $t = 0$, that is $P = |\langle \psi_b \mid \Psi \rangle|^2$.

Again, since the unperturbed eigenfunctions $\psi_k^{(0)}$ form a complete set of basis vectors, even a general solution to the Schrödinger equation of the total Hamiltonian can be expressed by them as

$$\Psi = \sum_k c_k(t) \psi_k^{(0)} \exp\left({}^{-iE_k^{(0)} t}/_{\hbar} \right) .$$

One finds that the probability amplitude for the system to be in the state ψ_b, for which the unperturbed state $\psi_b^{(0)}$ is a member of the set $\left\{ \psi_k^{(0)} \right\}$, can be expressed by

$$\frac{\partial}{\partial t} c_b(t) = \frac{\lambda}{i\hbar} \sum_k \langle \psi_b^{(0)} \mid \hat{H}' \mid \psi_k^{(0)} \rangle \exp\left(i\omega_{bk} t\right) c_k(t)$$

with the Born angular frequency

$$\omega_{bk} = \frac{E_b^{(0)} - E_k^{(0)}}{\hbar} .$$

Making use of the fact that the perturbation is small, the coefficients can be expanded in powers of λ

$$c_k = c_k^{(0)} + \lambda c_k^{(1)} + \lambda^2 c_k^{(2)} + \cdots$$

This in turn opens the possibility for the consecutive calculation of ${}^{d}/_{dt} c_b(t)$ starting from the known initial condition $c_k^{(0)}$:

$$\begin{aligned}
\frac{\partial}{\partial t} c_b^{(0)} =& 0 \\
\frac{\partial}{\partial t} c_b^{(1)} =& \frac{1}{i\hbar} \sum_k \langle \psi_b^{(0)} \mid \hat{H}' \mid \psi_k^{(0)} \rangle \exp\left(i\omega_{bk} t\right) c_k^{(0)} \\
\frac{\partial}{\partial t} c_b^{(s+1)} =& \frac{1}{i\hbar} \sum_k \langle \psi_b^{(0)} \mid \hat{H}' \mid \psi_k^{(0)} \rangle \exp\left(i\omega_{bk} t\right) c_k^{(s)}, \qquad s = 0, 1, \ldots .
\end{aligned}$$

With this series, to first order, the probability amplitude to find the electron in state ψ_b under the conditions that it was in state ψ_a for $t \leq 0$ and that the external field is switched on at time $t = 0$ is given by

$$c_b^{(1)}(t) = \frac{1}{i\hbar} \langle \psi_b \mid \hat{H}' \mid \psi_a \rangle \left(F(t, \omega_{ba} + \omega) - F(t, \omega_{ba} - \omega)\right) \tag{5.3.7}$$

where the last two functions contribute to the emission and absorption of energy respectively. These cases can now be evaluated separately, which leads to an expression for the absorption or emission of energy per electron from the field

$$(\hbar\omega_{ba})\, W_{ba} = \frac{\pi e^2 I(\omega_{ba})}{\hbar^2 c\varepsilon_0} \cos^2\theta\, |\langle \psi_b \mid \hat{r} \mid \psi_a \rangle|^2$$

where W_{ba} is the transition rate, that is the time derivative of the transition probability, $I(\omega_{ba})$ comes from 5.3.3, and θ is the angle between the polarisation vector of the electromagnetic field $\hat{\epsilon}$ and the position operator $\hat{r}$ of the atom. This term is the same for absorption and emission, which means that the system is in equilibrium.

I want to close this case study with brief comments on selection rules and energy conservation. First, one can show that only transitions from ψ_a to ψ_b are possible that fulfill certain conditions, called selection rules. For the present process, these rules are that the azimuthal quantum numbers l_a and l_b of both states have to differ by ± 1 and that the magnetic quantum numbers m_a and m_b have to be either the same or differ by ± 1. Finally, it has been said that, since the Hamiltonian of the electron is time-dependent, energy will not be conserved. On the other hand, it is of course true that energy is conserved for the entire system of the helium atom plus the radiation field. This is reflected by the fact that the first term in the bracket of 5.3.7, describing the contribution of emission, corresponds to an outgoing wave

$$\vec{A}(t) = A_0(\omega)\hat{\epsilon}\cos(\omega t)$$

leading to a state ψ_b with energy $E_b = E_a - \hbar\omega$. The second term corresponds to an incoming wave

$$\vec{A}(t) = A_0(\omega)\hat{\epsilon}\cos(-\omega t)$$

leading to a state ψ_b with energy $E_b = E_a + \hbar\omega$ (cf. Schulten 2000, p. 243). Thus, every change in the energy of the electron is accompanied by either an ingoing or an outgoing electromagnetic wave. A special case worth mentioning is that in which the perturbation does not depend on time. Without going into details, what one finds is that transitions for which $E_a \simeq E_b$, that is, both energies deviate only by a margin smaller than that of the time-energy uncertainty relation $\Delta\mathrm{E}\Delta\mathrm{t} \gtrsim \hbar$, can occur. Only when $t \to \infty$ it will be the case that $E_a = E_b$ (cf. Bransden & Joachain 2000, p. 436 f.).

5.3.1.2 Case study 2: Non-relativistic quantum scattering

Sub-atomic structure is most commonly investigated through scattering experiments and the description of these experiments is the primary use of QFT. This case study will present some of the basic concepts of non-relativistic scattering, which in the next section can easily be transferred into the full relativistic treatment of scattering.[9]

For the elastic scattering of two spinless particles of mass m_1 and m_2 the complete system can be described by the product of the wave functions of the two particles

$$\psi(\vec{r}_1, \vec{r}_2, t) = \psi_1(\vec{r}_1, t)\psi_2(\vec{r}_2, t)\,. \tag{5.3.8}$$

However, if the interaction only depends on the distance $\vec{r} = |\vec{r}_1 - \vec{r}_2|$ of the particles, we can introduce the reduced mass $\mu = {m_1 m_2}/{(m_1+m_2)}$, which reduces 5.3.8 to $\psi(\vec{r}, t)$. Furthermore, the interaction can be approximated by a time-independent potential $V(\vec{r})$, and if the incident beam is switched on for a very long time compared to the time of the interaction, it reduces to the stationary wave function given by $\psi(\vec{r}, t) = \psi(\vec{r})e^{-iEt/\hbar}$, which solves the Schrödinger equation

$$\left(-\frac{\hbar^2}{2\mu}\Delta + V(\vec{r})\right)\psi(\vec{r}) = E\psi(\vec{r})\,. \tag{5.3.9}$$

In experiments the incident wave is fairly well collimated, so that it does not interact with the detector before the scattering, but it has to be spatially large enough to keep the spreading of the wave small. However, if the interaction between the scattering particles is weak and only has an effect in a small region around the target, the incident wave can be approximated by a plane wave. This is the Born approximation:

$$\psi_{in}(\vec{r}) = e^{i\vec{k}_0\cdot\vec{r}}\,.$$

The outgoing wave after the scattering, on the other and, has the form of a spherical wave

$$\psi_{sc}(\vec{r}) = f(\theta, \varphi)\frac{e^{i\vec{k}\cdot\vec{r}}}{r}$$

with an amplitude $f(\theta, \varphi)$, dependent on the direction (θ, φ), which is called the scattering amplitude. After the scattering event, the total wave function is therefore given by

$$\psi(\vec{r}) = e^{i\vec{k}_0\cdot\vec{r}} + f(\theta, \varphi)\frac{e^{i\vec{k}\cdot\vec{r}}}{r}\,. \tag{5.3.10}$$

The probability current density of the wave function is given by the expression

$$\vec{j}(\vec{r}) = \frac{\hbar}{2\mu i}\left[\psi^*\left(\vec{\nabla}\psi\right) - \left(\vec{\nabla}\psi^*\right)\psi\right]$$

to which for large r only the radial component contributes, which is for the incident and scattered wave respectively

[9]The following exposition is based on Bransden & Joachain (2000, ch. 12), Greiner & Reinhardt (2009, ch. 1) and Zettili (2009, ch. 11).

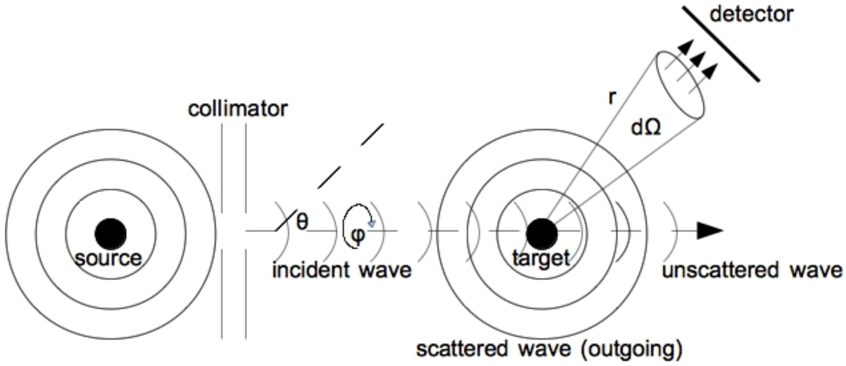

Figure 5.1: Schematic illustration of a scattering experiment

$$\begin{aligned} j_{in}^{(r)} &= \frac{\hbar \vec{k}_0}{\mu} \\ j_{sc}^{(r)} &= \frac{\hbar \vec{k}}{\mu} \frac{|f(\theta,\varphi)|^2}{r^2}\,. \end{aligned} \tag{5.3.11}$$

Now, an area of a detector has the size of $r^2 \mathrm{d}\Omega$ and therefore the quantity

$$\mathrm{d}N = j_r r^2 \mathrm{d}\Omega = \frac{\hbar \vec{k}}{\mu} |f(\theta,\varphi)|^2 \,\mathrm{d}\Omega \tag{5.3.12}$$

can be interpreted as the number of particles N hitting an area of the detector per unit time. Furthermore, conjoining 5.3.11 with 5.3.12 and taking into account that in the case of elastic scattering $\vec{k}_0 = \vec{k}$, one can define the scattering cross section

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = |f(\theta,\varphi)|^2$$

as the ratio of the number of scattered particles per unit area per unit time to the incident particle current density.

In order to find a solution to the scattering problem, the scattering amplitude can now be obtained from an integral formulation of the Schrödinger equation 5.3.9. First, since the wave function is stationary, the energy is given by $E = \hbar^2\vec{k}^2/2\mu$ and with the reduced potential $U(\vec{r}) = 2\mu/\hbar^2 V(\vec{r})$ 5.3.9 can be reformulated as

$$\left(\Delta + \vec{k}^2\right) \psi(\vec{r}) = U(\vec{r})\psi(\vec{r})\,. \tag{5.3.13}$$

A general solution to 5.3.13 consists of two parts, namely, a particular solution to the homogeneous part, which is just the incident wave, and a general solution, which can be found with the help Green's function $G_0(\vec{r} - \vec{r}')$. The general solution then is found to be

$$\psi(\vec{r}) = \psi_{in}(\vec{r}) + \int G_0(\vec{r} - \vec{r}')U(\vec{r}')\psi(\vec{r}')d^3\vec{r}' \tag{5.3.14}$$

where the integral extends over the whole space. The role of the Green's function $G(\vec{r} - \vec{r}')$, also called the propagator, can be understood from the following. In general, the Green's function describes how one wave $\psi(\vec{r}')$ relates to another wave $\psi(\vec{r})$, that is, for a free particle it will simply describe the evolution according to the free Schrödinger equation and different states will be related by the free Green's function

$$\psi(\vec{r}) = \int G_0(\vec{r} - \vec{r}')\psi(\vec{r}')\mathrm{d}^3x\,.$$

Since it is assumed that for the scattering experiment the motion of the free particle is known completely, the question is whether the Green's function for the interacting particles, $G(\vec{r} - \vec{r}')$, can be expressed by the free Green's function. It turns out that as a good approximation this is possible as

$$\psi_{sc}(\vec{r}) = \int G_0(\vec{r} - \vec{r}')U(\vec{r}')\psi(\vec{r}')\mathrm{d}^3x$$

and thus the general solution to the scattering Schrödinger equation is given by 5.3.14.

The next step in finding the Green's function explicitly is to note that the free Green's function must be a solution to the homogeneous, that is, the free part of 5.3.13

$$\left(\Delta + \vec{k}^2\right)\int G_0(\vec{r} - \vec{r}')\psi(\vec{r}')\mathrm{d}^3x = 0\,.$$

One finds that this can be reformulated as (cf. Greiner & Reinhardt 2009, p. 25)

$$\left(\Delta + \vec{k}^2\right)G_0(\vec{r} - \vec{r}') = \delta(\vec{r} - \vec{r}')$$

This equation has two solutions

$$\begin{aligned} G_0^+(\vec{r} - \vec{r}') &= -\frac{1}{4\pi}\frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r} - \vec{r}'|} \\ G_0^-(\vec{r} - \vec{r}') &= -\frac{1}{4\pi}\frac{e^{-ik|\vec{r}-\vec{r}'|}}{|\vec{r} - \vec{r}'|} \end{aligned}$$

corresponding to an outgoing wave from $\vec{r}$ and an incoming wave into $\vec{r}$ respectively. Since in the context of a scattering experiment only the outgoing wave is interesting, the general solution to 5.3.13 is now given by

$$\psi(\vec{r}) = \psi_{in}(\vec{r}) - \frac{1}{4\pi}\int \frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r} - \vec{r}'|}U(\vec{r}')\psi(\vec{r}')d^3\vec{r}'\,. \tag{5.3.15}$$

An explicit solution to this can be found perturbatively with the Born series. Here the zero-order solution is given simply by the incident wave, which can then be inserted in the first-order solution and so on

$$\begin{aligned}\psi_0(\vec{r}) =& e^{i\vec{k}_0\cdot\vec{r}}\\ \psi_1(\vec{r}) =& \psi_{in}(\vec{r}) - \frac{1}{4\pi}\int \frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}U(\vec{r}')\psi_0(\vec{r}')d^3\vec{r}'\\ \psi_n(\vec{r}) =& \psi_{in}(\vec{r}) - \frac{1}{4\pi}\int \frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}U(\vec{r}')\psi_{n-1}(\vec{r}')d^3\vec{r}'\end{aligned}$$

Furthermore, in the asymptotic limit of $r \to \infty$ 5.3.15 reduces to

$$\psi(\vec{r}) = e^{i\vec{k}\vec{r}} - \frac{1}{4\pi}\frac{e^{ikr}}{r}\int e^{-i\vec{k}\vec{r}'}U(\vec{r}')\psi(\vec{r}')d^3\vec{r}' \,. \tag{5.3.16}$$

Upon comparison of 5.3.10 and 5.3.16 one can now see that the scattering amplitude in an integral formulation is given by

$$f(\theta,\varphi) = -\frac{1}{4\pi}\int e^{-i\vec{k}\vec{r}'}U(\vec{r}')\psi(\vec{r}')d^3\vec{r}' = -\frac{1}{4\pi}\langle\psi_{in} \mid \hat{U} \mid \psi\rangle \,. \tag{5.3.17}$$

Here the matrix element $\langle\psi_{in} \mid \hat{U} \mid \psi\rangle$ is usually called an element of the scattering or S-matrix, which will receive more attention later. Generally, an element of the S-matrix S_{fi} gives the probability for the evolution of a state ψ_i into another state ψ_f.

In the first Born approximation, the scattering cross section is then

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = |f(\theta,\varphi)|^2 = \frac{1}{16\pi^2}\left|\int e^{-i\vec{k}\vec{r}'}U(\vec{r}')\psi_{in}(\vec{r}')d^3\vec{r}'\right|^2 = \frac{1}{16\pi^2}\left|\int e^{i\vec{q}\vec{r}'}U(\vec{r}')d^3\vec{r}'\right|^2$$

with $\hbar\vec{q} = \hbar\left(\vec{k}_0 - \vec{k}\right)$ as the momentum transfer from the incident to the scattered state.

An interesting special case to end this section is that of the scattering of identical particles. I will present here only the simplest case of identical bosons. What makes scattering of identical particles special is that there is no way to distinguish between the two possible final states, a) where particle 1 is scattered in direction (θ,φ) and particle 2 in the opposite direction $(\pi-\theta,\varphi+\pi)$, and case b) where particle 2 is scattered in direction (θ,φ) and particle 1 in $(\pi-\theta,\varphi+\pi)$. Therefore, both possibilities have to be taken into account when calculating the scattering cross section.

In classical physics one would expect the cross section to be simply the sum of both cases

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}_{cl} = |f(\theta,\varphi)|^2 + |f(\pi-\theta,\varphi+\pi)|^2$$

However, in quantum physics the final state after scattering is a superposition of both cases and consequently the cross section will be

$$\begin{aligned}\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}_{qm} &= |f(\theta,\varphi) + f(\pi-\theta,\varphi+\pi)|^2 \\ &= |f(\theta,\varphi)|^2 + |f(\pi-\theta,\varphi+\pi)|^2 + 2\mathrm{Re}\left[(\theta,\varphi)f^*(\pi-\theta,\varphi+\pi)\right]\end{aligned}$$

One can see that the quantum physical scattering can clearly be distinguished from the classical case by an additional interference term.

5.3.1.3 Case study 3: Quantum electrodynamics

Classical field theory[10] To begin with, it will be useful to start with the basics of classical field theory. In particular, this should help to clarify the role of the Lagrangian in QFT. A field is a function that assigns a certain variable $\phi(\vec{x},t)$ to every point in spacetime and, in the case of QFT, transforms under the Poincaré group. Consequently, the Lagrange function, $L = T - V$, becomes a functional of the field and its first time derivative

$$L(t) = L\left[\phi(\vec{r},t), \dot{\phi}(\vec{r},t)\right] .$$

Furthermore, since all Lagrange functions, appearing in fundamental physical theories today, are local, they can be expressed by a Lagrangian density[11]

$$L(t) = \int \mathcal{L}(\phi, \partial_\mu\phi)\mathrm{d}^3\mathbf{x} \tag{5.3.18}$$

which will henceforth be simply called the Lagrangian. This Lagrangian is local in so far that it only depends on one spatial vector $\mathbf{x}$ and not on any $\mathbf{y} \neq \mathbf{x}$. Nevertheless, the Lagrangian is in a certain sense not local in that it depends at every instant in time on the field values $\phi(\mathbf{x})$ over all of space.

The action integral is then given by

$$S = \int L(t)\mathrm{d}t = \int \mathcal{L}(\phi, \partial_\mu\phi)\mathrm{d}^4x$$

and from Hamilton's principle of the least action follow the Euler-Lagrange equations as

$$\frac{\partial\mathcal{L}}{\partial\phi_i} - \partial_\mu\frac{\partial\mathcal{L}}{\partial_\mu\phi_i} = 0 .$$

[10] The following exposition is based on Greiner & Reinhardt (1996, ch. 2) and Maggiore (2005, ch. 3).

[11] Here the usual relativistic notation will be used, in which $\partial_\mu = \partial/\partial x_\mu$ with spacetime coordinate x and the signature metric $\eta_{\mu\nu} = (+,-,-,-)$. $\Box = \partial^\mu\partial_\mu$ is the D'Alambert operator. The spatial three-vector will be denoted by $\mathbf{x}$. Natural units will be used, that is, $\hbar = c = 1$.

The Euler-Lagrange equations are the equations of motion for the fields ϕ_i. The most common special cases of them are the Klein-Gordon equation for scalar particles and the Dirac equation for spin-$1/2$ particles. Hence, for a free real scalar field with mass m the Lagrangian is

$$\mathcal{L}(x) = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2$$

from which Hamilton's principle leads to the Klein-Gordon equation

$$\left(\Box + m^2\right)\phi = 0\,. \tag{5.3.19}$$

A simple solution to 5.3.19 is a plane wave $e^{\pm ipx}$ with $p^2 = m^2$.

The Dirac equation on the other hand is the equation of motion for spin-$1/2$ particles and therefore states are represented by spinors, that is, fields with four complex components satisfying certain transformation laws. However, for simplicity the spinor indices will be omitted. Using the Dirac or gamma matrices γ^μ, the notation of the Feynman slash $\not{\partial} = \gamma^\mu\partial_\mu$ and the Dirac adjoint $\bar{\Psi} = \Psi^\dagger\gamma^0$ the Dirac Lagrangian can be written as

$$\mathcal{L}_{Dirac} = \bar{\Psi}\left(i\not{\partial} - m\right)\Psi \tag{5.3.20}$$

from which in the usual way the Dirac equation follows as

$$\left(i\not{\partial} - m\right)\Psi = 0\,.$$

Also useful for what is to come will be to take a look at the explicitly covariant formulation of electrodynamics. The electrodynamic field can be described by using a four-vector potential $A^\mu = \left(A^0, \vec{A}\right)$, which is related to the electric and the magnetic field strengths via

$$\begin{aligned}\vec{E} =& -\frac{\partial\vec{A}}{\partial t} - \vec{\nabla}A_0\\ \vec{B} =& \vec{\nabla}\times\vec{A}\,.\end{aligned}$$

Now, the field strength tensor can be defined as

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu\,.$$

The Lagrangian for the free electromagnetic field is then given by

$$\mathcal{L}_{e.m.} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} = \frac{1}{2}\left(\vec{E}^2 - \vec{B}^2\right) \tag{5.3.21}$$

From which follows the equation of motion

$$\partial_\mu F^{\mu\nu} = 0\,. \tag{5.3.22}$$

A coupling of the electromagnetic field to an external source or charged matter field can be reached by introducing the conserved vector current $j^\mu = q\bar{\Psi}\gamma^\mu\Psi$ of a Dirac field with charge q into the equation of motion 5.3.22 such that

$$\partial_\mu F^{\mu\nu} = j^\mu \tag{5.3.23}$$

which is a consequence of the Lagrangian

$$\mathcal{L} = \mathcal{L}_{e.m.} + \mathcal{L}_{int.} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - q\bar{\Psi}\gamma^\mu\Psi A_\mu\,. \tag{5.3.24}$$

As a final comment, the transition from the Lagrange to the Hamilton formalism is made by introducing the canonically conjugate field

$$\pi(x) = \frac{\partial\mathcal{L}}{\partial(\partial_0\phi_i)}\,.$$

A Legendre transformation of the Lagrangian then leads to the Hamiltonian density

$$\mathcal{H}(x) = \pi_i(x)\partial_0\phi_i(x) - \mathcal{L}$$

from which the total Hamiltonian follows as

$$H = \int \mathcal{H}\mathrm{d}^3x$$

Furthermore, the equations of motion are now given as

$$\begin{aligned}\dot{\phi} =& \frac{\partial\mathcal{H}}{\partial\pi} - \nabla\frac{\partial\mathcal{H}}{\partial(\nabla\pi)}\\ \dot{\pi} =& \frac{\partial\mathcal{H}}{\partial\phi} - \nabla\frac{\partial\mathcal{H}}{\partial(\nabla\phi)}\,.\end{aligned}$$

Canonical quantisation formalism We will now take the leap from classical field theory to QFT. Field quantisation is an involved topic (cf. Greiner & Reinhardt 1996), that here has to be omitted completely for the sake of brevity. I will simply take it as a given result of the quantisation procedure that classical fields can be promoted to quantum fields in the following way.

A general classical solution to the Dirac equation can be written as

$$\Psi(x) = \int \frac{\mathrm{d}^3p}{(2\pi)^3\sqrt{2E_p}} \sum_{s=1,2} \left(a_{p,s}u^s(p)e^{-ipx} + b^*_{p,s}v^s(p)e^{ipx}\right) \tag{5.3.25}$$

that is, as an expansion in a complete set of plane waves, with s being the degree of freedom of spin. This field is turned into a quantum field by imposing the equal time anticommutator relations

$$\left\{\hat{\Psi}_a(\mathbf{x},t),\hat{\Psi}_b^\dagger(y,t)\right\} =\delta_{ab}\delta^{(3)}(\mathbf{x}-\mathbf{y})$$
$$\left\{\hat{\Psi}_a(\mathbf{x},t),\hat{\Psi}_b(y,t)\right\} =\left\{\hat{\Psi}_a^\dagger(\mathbf{x},t),\hat{\Psi}_b^\dagger(y,t)\right\}=0\,.$$

Here the anticommutator ensures that the described Fermions, that is, spin-$1/2$ particles conform to Fermi-Dirac statistics, as opposed to Bose-Einstein statistics for Bosons. Furthermore, the coefficients $a_{p,s}$ and $b_{p,s}$ are promoted to operators, satisfying the equal time anticommutator relations

$$\left\{\hat{a}(p,s),\hat{a}^\dagger(q,r)\right\} =\left\{\hat{b}(p,s),\hat{b}^\dagger(q,r)\right\}=(2\pi)\delta_{sr}\delta^{(3)}(p-q)$$
$$\left\{\hat{a}(p,s),\hat{a}(q,r)\right\} =\left\{\hat{a}^\dagger(p,s),\hat{a}^\dagger(q,r)\right\}=0$$
$$\left\{\hat{b}(p,s),\hat{b}(q,r)\right\} =\left\{\hat{b}^\dagger(p,s),\hat{b}^\dagger(q,r)\right\}=0\,.$$

The operator $\hat{a}^\dagger(p,s)$ is the creation operator for particles with $U(1)$ charge $Q=1$, while $\hat{b}^\dagger(p,s)$ is the creation operator for antiparticles with charge $Q=-1$. The operators $\hat{a}(p,s)$ and $\hat{b}(p,s)$ are the annihilation operators respectively.

The vacuum state can now be defined as

$$\hat{a}(p,s)\left|0\right\rangle=\hat{b}(p,s)\left|0\right\rangle=0\,.$$

One can then use the creation operators to build up any state in Fock space from the vacuum state

$$\left|p_{1,s},\dots p_{n,s}\right\rangle=(2E_{p_1})^{1/2}\cdots(2E_{p_n})^{1/2}\,\hat{a}^\dagger(p_1,s)\cdots\hat{a}^\dagger(p_n,s)\left|0\right\rangle \tag{5.3.26}$$

with normalisation constants $(2E_{p_n})^{1/2}$.

The quantisation of the electromagnetic field on the other hand differs from that of the Dirac field, since the electromagnetic potential, being a gauge field, has more degrees of freedom. Furthermore, the fact that the photon has zero mass poses particular problems. However, the process can be simplified by working in the Coulomb gauge again, in which $\vec{\nabla}\cdot\vec{A}=0$ and $A_0=\phi=0$. A general solution to the classical wave equation for the electromagnetic potential 5.3.1 can be written as

$$\vec{A}=\int\frac{\mathrm{d}^3p}{(2\pi)^3\sqrt{2\omega_{\vec{p}}}}\sum_{\lambda=1,2}\left[\vec{\epsilon}(\vec{p},\lambda)a_{\vec{p},\lambda}e^{-ipx}+\vec{\epsilon}^*(\vec{p},\lambda)a^*_{\vec{p},\lambda}e^{ipx}\right]$$

with the polarisation vectors $\vec{\epsilon}(\vec{p},\lambda)$. This field is now quantised, by again promoting the coefficients $a_{\vec{p},\lambda}$ and $a^*_{\vec{p},\lambda}$ to operators, such that the operator field now reads

$$\hat{A}(x)=\int\frac{\mathrm{d}^3p}{(2\pi)^3\sqrt{2\omega_{\vec{p}}}}\sum_{\lambda=1,2}\left[\vec{\epsilon}(\vec{p},\lambda)\hat{a}_{\vec{p},\lambda}e^{-ipx}+\vec{\epsilon}^*(\vec{p},\lambda)\hat{a}^\dagger_{\vec{p},\lambda}e^{ipx}\right]\,.$$

The annihilation and creation operators $\hat{a}_{\vec{p},\lambda}$ and $\hat{a}^{\dagger}_{\vec{p},\lambda}$ now fulfil the equal time commutator relations

$$\begin{aligned}\left[\hat{a}_{\vec{p},\lambda}, \hat{a}^{\dagger}_{\vec{q},\lambda'}\right] =&(2\pi)^3\delta^{(3)}\left(\vec{p}-\vec{q}\right)\delta_{\lambda\lambda'}\\ \left[\hat{a}_{\vec{p},\lambda}, \hat{a}_{\vec{q},\lambda'}\right] =& \left[\hat{a}^{\dagger}_{\vec{p},\lambda}, \hat{a}^{\dagger}_{\vec{q},\lambda'}\right] = 0\,.\end{aligned} \tag{5.3.27}$$

Interactions After the quantisation of the Dirac and the electromagnetic fields is done, we are now in a position to describe interactions between fermions involving the electromagnetic force. A good example for an electromagnetic interaction is the so-called Bhabha scattering, that is, the scattering of electrons with their antiparticles, the positron, through which muons and anti-muons are produced. Up to energies of 91.16 GeV, the mass of the Z Boson, this interaction can be treated entirely in quantum electrodynamics and any influence from the electroweak force can be neglected (cf. Greiner & Reinhardt 2009, p. 142). Similar to the Hamiltonian in the first case study, the Lagrangian now consists of parts for the free motion of the matter and force fields and one term that couples both

$$\hat{\mathcal{L}}_{QED} =\hat{\mathcal{L}}_{Dirac} + \hat{\mathcal{L}}_{e.m.} + \hat{\mathcal{L}}_{int.} \tag{5.3.28}$$

$$\hat{\mathcal{L}}_{Dirac} =\hat{\bar{\Psi}}\left(i\partial\!\!\!/ - m\right)\hat{\Psi} \tag{5.3.29}$$

$$\hat{\mathcal{L}}_{e.m.} = -\frac{1}{4}\hat{F}_{\mu\nu}\hat{F}^{\mu\nu} - \frac{1}{2}\left(\partial_\mu\hat{A}^\mu\right)^2 \tag{5.3.30}$$

$$\hat{\mathcal{L}}_{int.} = -q\hat{\bar{\Psi}}\gamma^\mu\hat{\Psi}\hat{A}_\mu\,. \tag{5.3.31}$$

Most of the terms in this Lagrangian are already familiar from the presentation of classical field theory above. Only the second term on the right hand side of $\hat{\mathcal{L}}_{e.m.}$ is new and appears, because we are no longer working in the Coulomb gauge, but in the Feynman gauge. In what follows, it will be practical to use the Hamiltonian, derived from $\hat{\mathcal{L}}_{QED}$, which can also be divided into a free and an interaction part

$$\hat{H} = \hat{H}_0 + \hat{H}_{int.}\,. \tag{5.3.32}$$

From our discussion of non-relativistic scattering above, we already know that the final goal is to calculate the cross section, which can be derived in a straightforward way from the S-matrix element

$$S = \langle\psi_{out}|\hat{S}|\psi_{in}\rangle\ . \tag{5.3.33}$$

Here again the operator $\hat{S}$ is simply the time evolution operator, which can be written as

$$\hat{S} = \lim_{t_2\to+\infty}\lim_{t_1\to-\infty} e^{-i\hat{H}(t_2-t_1)}$$

Furthermore, $\hat{S}$ is a unitary operator. This can be seen if we take S to be the probability amplitude, a complex number, of which the absolute square gives the probability P for the initial state to evolve into the final state

$$P = |\langle\psi_{out}|\hat{S}|\psi_{in}\rangle|^2 \quad , \quad P \in \mathbb{R}|0 \leq P \leq 1$$

If we now sum up over all possible outcomes ψ_n, the probability is always unity

$$P = 1 = \sum_n |\langle\psi_n|\hat{S}|\psi_{in}\rangle|^2$$

which means that $\hat{S}$ must be unitary operator.

The states, on the other hand, are usually constructed as momentum and spin eigenstates, long before and after the scattering event. It is most common in QFT to work in the Heisenberg picture, in which states are constant

$$|\psi, t\rangle^H = |\psi, 0\rangle^H = |\psi\rangle^H$$

and operator fields evolve in time

$$\hat{\phi}^H(t) = e^{i\hat{H}t}\hat{\phi}^H(0)e^{-i\hat{H}t}$$

However, it is always possible to transfer into the Schrödinger picture, in which states evolve in time and operator fields are constant

$$\begin{aligned} |\psi, t\rangle^S =& e^{-i\hat{H}t}|\psi\rangle^H \\ \hat{\phi}^S =& e^{-i\hat{H}t}\hat{\phi}^H(t)e^{i\hat{H}t} . \end{aligned}$$

One of the basic methods to evaluate the S-matrix is to apply the LSZ reduction formula, named after Harry Lehmann, Kurt Symanzik and Wolfhart Zimmermann, who came to realise that the S-matrix 5.3.33 can be reformulated in a way that reduces its calculation essentially to the evaluation of vacuum expectation values for time ordered products of field operators

$$\langle 0|T\left\{\hat{\phi}(x_1)\ldots\hat{\phi}(x_n)\right\}|0\rangle \tag{5.3.34}$$

where the T indicates time ordering, viz., $t_1 > t_2 > \ldots > t_n$. This is what is called the n-point Green's function, $G(x_1, \ldots, x_n)$, the role of which can be elucidated by remembering how it was used in the previous case study of non-relativistic scattering. Given that with 5.3.26 any state can be build up from the vacuum, the possibility of the LSZ reduction should be intuitively clear. Now, keeping in mind that $\hat{S}$ is the time evolution operator, evaluating the Green's function amounts to calculating the time evolution under the condition that all momenta of the initial and final states are different, and means that all non-interacting particles are neglected.

The next problem one encounters is that interacting fields follow complicated equations of motion, derived from the full Lagrangian, which in general cannot be solved exactly. Thus, we do not know how exactly interacting fields look like. The workaround strategy is, similar to the case of non-relativistic scattering, to describe the interacting fields via a perturbation series of free fields. The first step is to change into the interaction picture, in which the interacting field $\hat{\phi}(x)$ can be expressed by the free interaction picture field $\hat{\phi}_I(x)$. Assuming that at one point in time t_0 both field are equal, $\hat{\phi}(\mathbf{x}, t_0) = \hat{\phi}_I(\mathbf{x}, t_0)$, they stand in the following relation

$$\hat{\phi}(\mathbf{x}, t) = e^{i\hat{H}(t-t_0)} e^{-i\hat{H}_0(t-t_0)} \hat{\phi}_I(\mathbf{x}, t) e^{i\hat{H}_0(t-t_0)} e^{-i\hat{H}(t-t_0)} .$$

Whereas the interaction picture Hamiltonian $\hat{H}_I$ is

$$\hat{H}_I(t) = e^{i\hat{H}_0(t-t_0)} \hat{H}_{int.} e^{-i\hat{H}_0(t-t_0)} .$$

It is then possible to show that the Green's function can be expressed as

$$\langle 0|T\left\{\hat{\phi}(x_1)\ldots\hat{\phi}(x_n)\right\}|0\rangle = \frac{\langle 0|T\left\{\hat{\phi}_I(x_1)\ldots\hat{\phi}_I(x_n)\exp\left[-i\int dt\hat{H}_I\right]\right\}|0\rangle}{\langle 0|T\left\{\exp\left[-i\int dt\hat{H}_I\right]\right\}|0\rangle} . \tag{5.3.35}$$

By expanding the exponential functions through a Taylor series, the Green's function can then be evaluated as a perturbation series in powers of $\hat{H}_I$. The use of perturbation theory is possible for QED in particular, since its coupling constant $\alpha = 1/137$ is relatively small.

Finally, the full S-matrix element for the example of QED can be found by conjoining the Green's functions for the Fermions, Photons and the interaction part up to the desired order.[12] The calculation of the cross section is then relatively easy.

Path integral formalism[13] An alternative route to the canonical quantisation method is the path integral method, which leads to equivalent results. Contrary to what we did before, in the path integral method fields are not promoted to operators, but remain classical. The basic task will then be to calculate a special kind of path integral over classical fields. Quantum physical phenomena, on the other hand, are replaced by the idea that one field configuration does not change into another configuration along only one path in phase space, as would be the case in classical mechanics, but along all possible paths at the same time. One of the advantages of the path integral method is that it does not necessarily involve perturbation theory. This is especially useful, since the evaluation of 5.3.35 in a perturbative expansion is not possible for all couplings, in particular when the weak and strong force become relevant, and thus the path integral method has to be used.

The general concepts of the path integral method are easiest to explain on a one-dimensional non-relativistic system. In the Heisenberg representation the system

[12]More on this will be said in the chapter about Feynman diagrams.
[13]This presentation is based on Greiner & Reinhardt (1996, ch. 11) and Maggiore (2005, ch. 9).

is described by the time dependent position and momentum operators $\hat{q}(t)$ and $\hat{p}(t)$ with corresponding eigenstates

$$\begin{aligned}\hat{q}(t)\,|q,t\rangle =& q\,|q,t\rangle \\ \hat{p}(t)\,|p,t\rangle =& p\,|p,t\rangle \ .\end{aligned}$$

For given t, the states $|p,t\rangle$ form a complete basis of the Hilbert space and are normalised plane waves, as shown by the inner product

$$\langle q|p\rangle = e^{ipq}\,.$$

As we already know from the discussion of canonical quantisation, the basic expression that needs to be computed is the transition amplitude, that is, the probability for an initial state to evolve into a final state

$$\langle q_f, t_f|q_i, t_i\rangle$$

which is often called the Feynman kernel. In evaluating this, we split up the time between the initial and the final state into equal intervals

$$t_n = t_0 + n\epsilon, \quad \text{with } t_f - t_i = N\epsilon$$

and $t_0 = t_i$, $t_N = t_f$. Since to each of these points in time there belongs a complete set of basis states, the Feynman kernel can be written as a direct product

$$\langle q_f, t_f|q_i, t_i\rangle = \int dq_1 dq_2 \ldots dq_{N-1} \prod_{n=0}^{N-1} \langle q_{n+1}, t_{n+1}|q_n, t_n\rangle \ . \tag{5.3.36}$$

Changing now into the Schrödinger representation, every intermediate Feynman kernel can be written as

$$\langle q_{n+1}, t_{n+1}|q_n, t_n\rangle = \langle q_{n+1}|e^{-i\hat{H}\epsilon}|q_n\rangle$$

and inserting a complete set of momentum eigenstates leads to

$$\begin{aligned}\langle q_{n+1}, t_{n+1}|q_n, t_n\rangle =& \int dp_n\, \langle q_{n+1}|p_n\rangle\, \langle p_n|e^{-i\hat{H}\epsilon}|q_n\rangle \\ =& \int dp_n e^{iq_{n+1}p_n}\, \langle p_n|e^{-i\hat{H}\epsilon}|q_n\rangle\end{aligned}$$

Given that the time intervals, that is, the ϵ are made sufficiently small, the exponential function can be approximated by the first two terms of a Taylor series

$$\langle q_{n+1}, t_{n+1}|q_n, t_n\rangle = \int dp_n e^{iq_{n+1}p_n}\, \langle p_n|1 - i\hat{H}(\hat{p},\hat{q})\epsilon + O(\epsilon^2)|q_n\rangle \ . \tag{5.3.37}$$

Since now the operators $\hat{p}$ and $\hat{q}$ act only on their respective eigenstates, we can switch from the quantum Hamiltonian $\hat{H}(\hat{p}, \hat{q})$ to the classical Hamiltonian $H(p, q)$ and rewrite 5.3.37 as

$$\begin{aligned}\langle q_{n+1}, t_{n+1}|q_n, t_n\rangle &= \int dp_n e^{iq_{n+1}p_n}\left(1 - i\hat{H}(\hat{p}, \hat{q})\epsilon + O(\epsilon^2)\right)\langle p_n|q_n\rangle \\ &= \int dp_n \exp\left(-i\epsilon\left[H(p_n, q_n) - p_n\frac{q_{n+1} - q_n}{\epsilon}\right]\right) + O(\epsilon^2)\,.\end{aligned}$$

Inserting this back into 5.3.36 leads to

$$\begin{aligned}\langle q_f, t_f|q_i, t_i\rangle = &\int dp_0(dq_1dp_1)\ldots(dq_{N-1}dp_{N-1}) \\ &\times \exp\left(-i\epsilon\sum_{n=0}^{N-1}\left[H(p_n, q_n) - p_n\frac{q_{n+1} - q_n}{\epsilon}\right]\right) + O(\epsilon^2)\,.\end{aligned} \tag{5.3.38}$$

The next step is now to let $\epsilon \to 0$ and consequently to regard q and p as functions of time with $q(t_n) = q_n$ and $p(t_n) = p_n$. We can then do the following substitutions in 5.3.38

$$\frac{q_{n+1} - q_n}{\epsilon} \to \dot{q}(t_n)\,, \quad \epsilon\sum_{n=0}^{N-1} f(t_n) \to \int_{t_i}^{t_f} dt\, f(t)\,.$$

Additionally, the following notation will denote path integrals

$$\mathcal{D}p\mathcal{D}q = \lim_{\epsilon\to 0} dp_o(dq_1dp_1\ldots(dq_{N-1}dp_{N-1})$$

with which 5.3.38 can be written as

$$\langle q_f, t_f|q_i, t_i\rangle = \int_{q(t_i)=q_i}^{q(t_f)=q_f}\mathcal{D}q\int\mathcal{D}p\, \exp\left(i\int_{t_i}^{t_f} dt\,[p\dot{q} - H(p, q)]\right)\,. \tag{5.3.39}$$

If now the Hamiltonian has the form of

$$H(p, q) = \frac{1}{2m}p^2 + V(q)$$

the integration over p can be performed explicitly and 5.3.39 becomes

$$\langle q_f, t_f|q_i, t_i\rangle = \int_{q(t_i)=q_i}^{q(t_f)=q_f}\mathcal{D}q\, \exp(iS) \tag{5.3.40}$$

with S being the classical action functional

$$S(q,\dot{q}) = \int_{t_i}^{t_f} dt \, \Big(\frac{m}{2}\dot{q}^2 - V(q)\Big) = \int_{t_i}^{t_f} dt \, L(q,\dot{q}) \, .$$

As a side note, this formulation suggests a particular, but not uncontroversial, interpretation due to Feynman (1985), and this is why 5.3.40 is sometimes called Feynman's path integral. Since the path integral runs over all possible functions $q(t)$ with the boundary conditions $q(t_i) = q_i$ and $q(t_f) = q_f$, and each function $q(t)$ describes a different path, according to Feynman, formula 5.3.40 might be interpreted as giving the probability amplitude for a particle to travel from q_i to q_f along all possible paths.

Transferring to field theory is now rather simple and can be done by replacing the function $q(t)$ by the scalar field $\phi(x)$, which also means that the Lagrangian L can be expressed via the Lagrangian density $\mathcal{L}$. It is important that the fields are now solutions to the equations of motion coming from the complete Hamiltonian 5.3.32, rather than only the Hamiltonian of the interaction picture. This means that the field $\phi(x)$ forms a set of eigenstates for the operator field $\hat{\phi}(x)$

$$\hat{\phi}(\mathbf{x},t)\,|\phi,t\rangle = \phi(\mathbf{x})\,|\phi,t\rangle$$

and the operator field satisfies the equation of motion

$$-i\dot{\hat{\phi}}(\mathbf{x},t) = \Big[\hat{H}, \hat{\phi}(\mathbf{x},t)\Big] \, .$$

Under the condition that the fields vanish in the infinite past and future, that is, $\phi_i(\mathbf{x}) = \phi_f(\mathbf{x}) = 0$ and $t_i \to -\infty$, $t_i \to -\infty$, one can show that vacuum to vacuum transition amplitude can be written as

$$\langle 0, t_f | 0, t_i \rangle = \int \mathcal{D}\phi \, e^{iS} \, . \tag{5.3.41}$$

Furthermore, the expectation value for a time-ordered product of fields can be written as

$$\langle 0, t_f | T\{\hat{\phi}(x_1) \ldots \hat{\phi}(x_n)\} | 0, t_i \rangle = \int \mathcal{D}\phi \, \hat{\phi}(x_1) \ldots \hat{\phi}(x_n) e^{iS}. \tag{5.3.42}$$

Conjoining the last two expressions, it follows that the vacuum expectation value for a time-ordered product of fields, that is, the n-point Green's function 5.3.35 is given as

$$\langle 0 | T\{\hat{\phi}(x_1) \ldots \hat{\phi}(x_n)\} | 0 \rangle = \frac{\int \mathcal{D}\phi \, \hat{\phi}(x_1) \ldots \hat{\phi}(x_n) e^{iS}}{\int \mathcal{D}\phi \, e^{iS}} \, . \tag{5.3.43}$$

We have therefore reached our goal to express the central term of scattering theory in the path integral formulation. So far, we have only discussed the case in which the fields satisfy canonical commutator relations. The generalisation to fields satisfying anticommutator relations will not be discussed here. Suffice it to say that this case can be described using classical fields and the help of anti-commuting Grassmann numbers (cf. Peskin & Schroeder 1995, p. 299 ff.).

This ends the presentation of how in general interactions are described in QFT. In the remaining sections of this chapter, I will take a closer look on particular features of this description that later may or may not serve as the basis for a theory of causation.

5.3.2 Group structure

The significance of group theory for physics was first presented in a systematic fashion by Wigner (1939). Though it should not be left unmentioned that important advances in group theory, and in particular Lie groups, were done by Weyl earlier (cf. French 1999, p. 194). Today, group theory is one of the fundamental tools, used to solve all kinds of problems in physics. In this section, I want to take a closer look on one specific element of group theory, namely, Wigner's classification. This is the idea that all elementary particles can be ordered into kinds by the group they belong to.[14] In order to come to an understanding of how this works, also in relation to the preceding paragraphs on the description of interactions, it will be useful to start with the basic mathematics of group theory.[15]

A group, G, is a set, $g_1, g_2, \ldots, g_n \in G$, together with an operation, $(\circ)$, which assigns to every pair of elements a third element, satisfying

1. if $g_i, g_j \in G \Rightarrow g_i \circ g_j = g_k \in G$
2. $g_i \circ (g_j \circ g_k) = (g_i \circ g_j) \circ g_k$
3. there is an identity element g_1, such that for all $g_i \in G$, $g_i \circ g_1 = g_1 \circ g_i = g_i$
4. for every element $g_k \in G$ there exists an inverse element $g_l \in G$, such that $g_k \circ g_l = g_l \circ g_k = g_1$.

A linear representation of G is a mapping D of the elements of G onto a set of linear operators

$$g_i \to D(g_i)$$

which act on a base space, such that

1. $D(g_1) = 1$, viz., there is an identity operator
2. $D(g_i)D(g_j) = D(g_i \circ g_j)$, which means that the group structure, given by the group multiplication, is preserved in the representation.

[14]To be precise, however, it is not entirely clear what Wigner thought to be classifying by group theory, since his terminology varies between 'state', 'system of equations' and 'particle'.

[15]Useful literature on group theory in physics, on which my presentation is based, is Gilmore (1974), Georgi (1999) and Maggiore (2005, ch. 2).

One important representation in physics is the matrix representation. The base space then is a linear vector space, with elements $(\phi_1, \phi_2, \ldots, \phi_n)$ and a group element g_k is represented by a $n \times n$ matrix $\left(D(g_k)\right)^i{}_j$. A representation of a group element then forms a transformation of the vector space

$$\phi^i \to \left(D(g_k)\right)^i{}_j \phi^j \,.$$

In QFT, however, since we are dealing with fields with infinite degrees of freedom, the matrix representation has to be generalised to a field representation.

A representation is called reducible, if it has an invariant subspace, such that for all ϕ^j, elements of the subspace, the transformed elements $\phi_i \to \left(D(g_k)\right)^i{}_j \phi^j$ belong to that subspace again. If this is not the case, the representation is called irreducible. It is common to use the term 'representation' for the base space as well as for elements of the base space and the representations of the group elements. I will follow this convention only, if it is clear from the context, which one of the latter is meant.

Additionally, important concepts are Lie groups and generators thereof. If the elements of a group depend smoothly on a set of continuous parameters

$$g(\theta^a)\,, \qquad a = 1, \ldots, N$$

the group is called a Lie group. The parameters θ^a will be chosen such that $\theta^a = 0$ corresponds to the identity element, $g(0) = g_1$. The same is true for the representation of the group, that is $D(\theta)|_{\theta=0} = 1$. Since the dependence on the parameters is smooth, every group element close to the identity element can be reached by the first two terms of a Taylor expansion

$$D(\theta) \simeq 1 + i\theta_a T^a$$

and

$$T^a \equiv -i \left. \frac{\partial D(\theta)}{\partial \theta_a} \right|_{\theta=0} \,. \tag{5.3.44}$$

The T^a are called the generators of the group. The i is added to 5.3.44 so that if the representation is unitary, the generators are hermitian operators. Under certain conditions, the representation of a group element $g(\theta)$ can be written as

$$D\left(g(\theta)\right) = e^{i\theta_a T^a}$$

Furthermore, the generators form an algebra, that is, they satisfy a commutator relation

$$\left[T^a, T^b\right] = i f^{ab}{}_c T^c$$

Here the $f^{ab}{}_c$ are called structure constants of the group. They are independent of the representation and define the Lie algebra.

If we now go into physics, one finds that the complete symmetry group of the standard model is the Poincaré group ISO$(3,1)$, that is the semidirect product of the proper Lorentz and the translation group

$$\mathrm{ISO}(3,1) = \mathrm{SO}(3,1) \rtimes \mathbb{R}^{3,1}\,.$$

To explicate this group in more detail, it is helpful to consider the Lorentz and the translation group separately.

The Lorentz group is the group of coordinate transformations

$$x^\mu \to x'^\mu = \Lambda^\mu_{\ \nu} x^\nu$$

that leave invariant the inner product on Minkowski space

$$\eta_{\mu\nu} x^\mu x^\nu = x_\nu x^\nu = t^2 - x^2 - y^2 - z^2\,.$$

An element of the Lorentz group can be written as

$$\Lambda_L = e^{-i\frac{1}{2}\omega_{\mu\nu}J^{\mu\nu}}\,.$$

where the parameters $\omega_{\mu\nu}$ split up in the parameters for rotations and boosts. The generators $J^{\mu\nu}$, on the other hand, satisfy the commutator relations, that is, the Lie algebra of SO(3,1)

$$[J^{\mu\nu}, J^{\rho\sigma}] = i\left(\eta^{\nu\rho}J^{\mu\sigma} - \eta^{\mu\sigma}J^{\nu\sigma} - \eta^{\nu\sigma}J^{\mu\rho} + \eta^{\mu\sigma}J^{\nu\rho}\right)\,. \tag{5.3.45}$$

If we now define the generators for rotations, J^i, and boosts, K^i

$$J^i = \frac{1}{2}\epsilon^{ijk}J^{ik}, \qquad K^i = J^{i0}$$

5.3.45 can be written as

$$\begin{aligned}\left[J^i, J^j\right] =& i\epsilon^{ijk}J^k\\ \left[J^i, K^j\right] =& i\epsilon^{ijk}K^k\\ \left[K^i, K^j\right] =& -i\epsilon^{ijk}J^k\,.\end{aligned} \tag{5.3.46}$$

Later on, the Casimir operators will play a central role. They are defined as the operators that commute with all the generators of a group. Most relevant for us is the case when the studied representation is irreducible, in which case the Casimir operators are proportional to the identity matrix. In the case of the Lorentz group, the Casimir operator is

$$J^2 = j(j+1)I_n \tag{5.3.47}$$

with I_n being the identity matrix and $j = 0, 1/2, 1, \ldots$.

By again defining the two generators

$$J^{+,i} = \frac{1}{2}\left(J^i + iK^i\right)$$
$$J^{-,i} = \frac{1}{2}\left(J^i - iK^i\right)$$

one finds that 5.3.46 can be written as

$$\begin{aligned}\left[J^{+,i}, J^{+,j}\right] &= i\epsilon^{ijk}J^{+,k} \\ \left[J^{-,i}, J^{-,j}\right] &= i\epsilon^{ijk}J^{-,k} \\ \left[J^{+,i}, J^{-,j}\right] &= 0\,.\end{aligned} \tag{5.3.48}$$

The first two lines are Lie algebras of the rotation group SU(2). Thus, to find out more about the Lorentz group, it is sufficient to analyse the rotation group SU(2), at least as long as one stays close to the identity element, since only there the algebras are valid and SO(3) and SU(2) are equivalent. Additionally, states, which transform under representations of SU(2), can be labelled by j_z with the possible values $-j, -j+1, \ldots, j$.

If we now turn to the other part of the Poincaré group, the translations, one finds that an element of the translation group can be written as

$$\Lambda_P = e^{-iP^\mu a_\mu}$$

where the generator P^μ is the momentum operator and the parameter a_μ gives the coordinate transformation $x^\mu \to x^\mu + a^\mu$. The generators satisfy the commutator

$$[P^\mu, P^\nu] = 0\,.$$

The Poincaré group has two Casimir operators. The first one is $P_\mu P^\mu$. If $|p\rangle$ is a one particle state with momentum $p = m$ in the rest frame, the eigenvalue of $P_\mu P^\mu$ is given by (cf. Ryder 2002, p. 60)

$$P_\mu P^\mu |p\rangle = m^2 |p\rangle\ .$$

The second Casimir operator is $W_\mu W^\mu$ with

$$W^\mu = -\frac{1}{2}\epsilon^{\mu\nu\rho\sigma}J_{\nu\rho}P_\sigma\,.$$

To find the eigenvalues of this operator, one has to distinguish between the massive and the massless case.

In the first case, one finds the eigenvalues

$$-W_\mu W^\mu = m^2 j(j+1), \qquad \text{for} \quad m \neq 0\,.$$

Massive representations can therefore be identified by the values of the rest-mass m and the spin $j = 0, 1/2, 1, \ldots$. In the massless case, on the other hand, $P^2 = 0$

and there is no rest frame. Instead, representations are labelled by the helicity h, which is the spin value in the direction of propagation of the particle. One finds that the possible values of the helicity are $h = 0, \pm 1/2, \pm 1, \ldots$, and helicity states transform under the group SO(2), a subgroup of the Poincaré group. The spin of massless particles thus no longer transforms under SU(2) and is in this respect different from the spin of a massive particle (cf. Ryder 2002, p. 63). Furthermore, states with positive and negative helicity are related by a parity transformation. One therefore does not refer to them as different particles, but as different helicity states of the same particle, or left and right handed states (cf. Maggiore 2005, p. 40). An additional complexity is necessary to distinguish photons from gluons, since both are massless and have a helicity of ± 1. Nevertheless, they can be distinguished by the colour charge. While the photon is colourless, the gluon has colour and belongs to the colour group $\mathrm{SU}(3)_{\mathrm{colour}}$ (cf. Sternberg 1995, ch. 5.12; Ryder 2002, ch. 1.12).

If we now return to Wigner's idea that kinds of particles can be distinguished by the labels of their respective irreducible group representations, the above discussion can be summarised as follows

$$\text{Poincaré group}\begin{cases} m \neq 0, \quad j = 0, 1/2, 1, \ldots \begin{cases} \text{electron} & m \approx 0.511\text{MeV}, j = 1/2 \\ \text{muon} & m \approx 105.7\text{MeV}, j = 1/2 \\ \vdots & \end{cases} \\ m = 0, \quad h = 0, \pm 1/2, \pm 1, \ldots \begin{cases} \text{photon} & h = \pm 1 \\ \text{gluon} & h = \pm 1, \ \text{colour charge} \\ \text{(graviton} & h = \pm 2) \end{cases} \end{cases}$$

Of course, whether a particle, corresponding to a certain mass and spin state, exists, is a question that can be answered by experiments only. Also, it should be mentioned that this classification is minimal in the sense that more properties could be included, such as parity, which however lead not to any further distinctions between particles.

To end this exposition of group structure in QFT, it will be instructive to give an example, which relates to the presentation of interactions in chapter 5.3.1.3 above (cf. Ryder 2002, p. 60; Duncan 2012, p. 129; Maciejko). In the discussion of the canonical quantisation method we encountered fermionic spinor fields, and found that according to 5.3.26 any state in Fock space can be build up from the vacuum space using creation operators. The one particle state with momentum p and spin s is therefore proportional to

$$|p, s\rangle = \hat{a}^\dagger(p, s) |0\rangle \ .$$

Now, if Λ is an element of a the Poincaré group and $D(\Lambda)$ a unitary representation thereof, the creation operator and the one particle state are constructed such that they transform as

$$D(\Lambda) |p, s\rangle \equiv |\Lambda p, s\rangle = \hat{a}^\dagger(\Lambda p, s) |0\rangle$$

and

$$D(\Lambda)\hat{a}^\dagger(p,s)D^{-1}(\Lambda) = \hat{a}^\dagger(\Lambda p, s)\,.$$

One sort of particle, which play a role in QED, are electrons that are described by Dirac fields of the form given in 5.3.25. Accordingly, the state and the creation operator are elements of the base space of a representation of the Poincaré group and the representations $D(\Lambda)$ are Dirac representations.

By a Dirac representation, the following is meant. We recall that different irreducible representations of SU(2) can be identified by the eigenvalues, j, of the Casimir operators with values $0, 1/2, 1, 3/2, \ldots$. The generator of boosts, K^i, on the other hand, does not correspond to a conserved quantity and is therefore not used to identify different representations. The special case in which $j = 1/2$ is called the spinor representation and is fundamental, because all other relevant representations can be build up from tensor products of spinors. Since according to 5.3.48 the Lorentz group is given by two algebras for J^+ and J^- respectively, the spinor representations of the Lorentz group can be labeled by two half-integers j_+ and j_-. Spinors with eigenvalues $j_+ = 1/2,\ j_- = 0$ or $j_+ = 0,\ j_- = 1/2$ are called left and right handed Weyl spinors respectively. A Dirac field Ψ_D can then be constructed from two Weyl spinors and thus has four components

$$\Psi_D = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}.$$

In summary, electrons are described by fields that are elements of the base space for massive spin $1/2$ representations of the Poincaré group. Or in other words, electrons are massive particles, with an empirically determined mass of $m \approx 0.511$MeV, and a spin of $1/2$. Other Fermions, which play a role in QED, will have a different mass value accordingly. Finally, since the electromagnetic field in the Lagrangian of QED 5.3.28 transforms under representations of U(1), its corresponding particle is massless and found to be the photon with helicity $h = \pm 1$.

Now, to return to the philosophical implications of group theory, I wish to argue that the rest mass and spin values indeed provide a sensible way to sort fundamental particles into kinds. This position seems to be so clear intuitively that hardly anyone gives reasons for it. Nevertheless, there have to be reasons why rest mass and spin are taken to be the relevant properties and not, say, the total energy of a system. Lyre (2012, p. 172) identifies two reasons, namely, "that physical systems must possess relativistically invariant state spaces with the most elementary, irreducible representations possessing no invariant subspaces." So on the one hand it is a reasonable requirement for elementary systems to be relativistically invariant, that is, to be symmetric under coordinate transformations and to 'look' the same for different observers in different inertial frames. This by itself excludes properties like the total energy from the definition, since the latter is not relativistically invariant. As shown above, the Poincaré group is sufficient to distinguish between particles, and other labels do not lead to any more fine graining, that is, more particles. On the other hand, the second reason is the irreducibility of the representations, which means that the representations cannot be divided further into more fundamental parts. In conclusion, the first reason shows that we are indeed dealing with mind independent entities and not just artefacts of the observer's perspective, while the second reason makes clear that this object is elementary in the sense that it cannot be split up further. It is clear, however, that these arguments do not explain why it is rest-mass and spin giving the distinctions into kinds and not another, relativistically invariant property. To this concern I want to answer that in the end it can be regarded as a brute fact of nature, in line with many other brute facts in physics. There is no explanation why rest-mass and spin characterise the kinds, they just do.[16]

5.3.3 Noether's theorem for field theories

In this section, I will first discuss Noether's theorem in classical field theories, before going into quantum physics. Noether's theorem connects every continuous transformation that leaves the action integral invariant with a conserved quantity (cf. Greiner & Reinhardt 1996, ch. 2.4; Maggiore 2005, ch. 3.2.1). To show this, it is sufficient here to study only infinitesimal transformations, which can be defined as follows

$$\begin{aligned} x^\mu \to & x'^\mu = x^\mu + A^\mu(x) \\ \phi(x) \to & \phi'(x') = \phi(x) + F(\phi, \partial\phi)\,. \end{aligned}$$

The transformations $A^\mu(x)$ and $F_i(\phi, \partial\phi)$ are defined as symmetry transformations, that is, they leave the action integral $S(\phi)$ invariant. This condition can be written as

$$\delta S(\phi) = \int\limits_V d^4x' \mathcal{L}'(x') - \int\limits_V d^4x \mathcal{L}(x) = 0\,.$$

By calculating this expression, one finds the following continuity equation

$$\frac{\partial}{\partial x_\mu} f_\mu(x) = 0 \tag{5.3.49}$$

with the current density

$$f_\mu(x) = \frac{\partial \mathcal{L}(x)}{\partial(\partial^\mu \phi)} \left(A^\nu(x)\partial_\nu \phi - F(\phi, \partial\phi)\right) - A^\mu(x)\mathcal{L}\,.$$

[16] Throughout this section I was only concerned with particles for which we have empirical evidence. These particles, however, are not the only ones allowed by group structure. For example, besides Fermions and Bosons, it is at least theoretically possible that there are paraparticles, following parastatistics (cf. French & Krause 2006, sec. 3.8). To regard something as a brute fact of nature is not meant as an argument to exclude the possibility of there being more than these brute facts or the brute facts to be wrong.

Integrating 5.3.49 leads to

$$0 = \frac{d}{dx_0} \int\limits_V d^3x \, f_0(x) + \oint\limits_{\partial V} d\vec{s} \, \vec{f}(x) \,. \tag{5.3.50}$$

If it is assumed that the field falls off to infinity sufficiently fast, then the integral over the surface ∂V is zero and

$$G := \int\limits_V d^3x \, f_0(x)$$

is constant in time, that is, it is a conserved quantity.

It will be important for my further discussion to note that the continuity equation 5.3.49 defines a local conservation law. Which means that either the quantity G is conserved in the volume V or it is not conserved, but then there is a current through the surface element ∂V. It follows that a conserved quantity cannot simply vanish at one point in space and appear at another distant point. A process of the latter sort is only possible with global conservation laws, but not with local conservation laws (cf. Ryder 2002, p. 79). That said, it has to be kept in mind that the conserved current is not observable in field theories. It is possible to construct two theories, which have the exact same equations of motion for the fields and therefore are empirically equivalent, but nevertheless have different conserved currents. Only the conserved quantity G in volume V is uniquely defined. As an additional consequence, the conserved quantity cannot in general be exactly localised at points, because V has to be chosen such that the field falls to zero at the boundaries and the surface integral in 5.3.50 does not contribute.[17]

An example for a conserved quantity is energy-momentum, which follows from symmetry under spacetime translations. A translation is given by the transformation

$$x^\mu \to x'^\mu = x^\mu + \epsilon^\mu$$

while the fields remain invariant, that is, $\phi'(x') = \phi(x)$. The conserved current is then the energy-momentum tensor

$$\theta_{\mu\nu} = \frac{\partial \mathcal{L}(x)}{\partial(\partial^\mu \phi)} \frac{\phi}{\partial^\nu} - \eta_{\mu\nu}\mathcal{L}$$

It follows that the conserved quantity is the energy-momentum four-vector

$$P^\nu = \int\limits_V d^3x \, \theta^{0\nu} = (E, \vec{p}) \ .$$

Going into QFT, the condition of energy-momentum conservation can be expressed as follows. As we know from 5.3.33 the basic expression one has to calculate is the

[17] Things might however be different if the gravitational field is included, which couples to the energy-momentum density (cf. Maggiore 2005, p. 69; Duncan 2012, p. 433).

matrix element for the evolution of an initial into a final state. With the evolution operator in the interaction picture, $\hat{\mathcal{H}}$, this can be expressed as

$$-i \int d^4x \, \langle \psi_{out} | \mathcal{H}(x) | \psi_{in} \rangle \ .$$

With the help of the operator $\hat{P}^\mu$ for spacetime translations, the evolution operator can be written as

$$\mathcal{H}(x) = e^{i\hat{P}x} \mathcal{H}(0) e^{-i\hat{P}x}$$

It follows that

$$\begin{aligned} -i \int d^4x \, \langle \psi_{out} | e^{i\hat{P}x} \mathcal{H}(0) e^{-i\hat{P}x} | \psi_{in} \rangle &= -\, i \int d^4x \, e^{i(P_{out} - P_{in})x} \, \langle \psi_{out} | H(0) | \psi_{in} \rangle \\ &= -\, i \, (2\pi)^4 \, \delta^{(4)} \, (P_{in} - P_{out}) \, \langle \psi_{out} | H(0) | \psi_{in} \rangle \end{aligned}$$

where the conservation is made explicit by the Dirac delta function (cf. Maggiore 2005, p. 205).

The influence of energy-momentum conservation also becomes obvious when the interaction between different fields is considered. In the discussion of QED, we have seen that the QED Lagrangian, 5.3.28, consists of three parts, namely, of the matter field the electromagnetic field and the interaction field. It can be shown that the continuity equation, 5.3.49, for the whole system then also has three parts, which are the energy-momentum tensors for the three fields

$$\partial_\mu \theta^{\mu\nu}_{Dirac} + \partial_\mu \theta^{\mu\nu}_{e.m.} + \partial_\mu \theta^{\mu\nu}_{int.} = 0 \, .$$

It follows that energy-momentum is conserved for the whole system and if, e.g., the energy-momentum of the Dirac field changes, it has to change for the other fields accordingly (cf. Greiner & Reinhardt 1996, exercise 6.2).

Finally, an example for another conserved quantity is the norm of the Dirac field 5.3.25, which follows from the symmetry under global phase transformations, so-called internal symmetries. Phase transformations can be expressed as $\psi \to \psi e^{i\chi}$ and $\psi^\dagger \to \psi^\dagger e^{-i\chi}$. The corresponding current destiny is then

$$j_\mu = -i \left(\frac{\partial \mathcal{L}(x)}{\partial(\partial^\mu \psi)} \psi - \frac{\partial \mathcal{L}(x)}{\partial(\partial^\mu \psi^\dagger)} \psi^\dagger \right) = \bar{\psi} \gamma_\mu \psi$$

and the conserved charge is

$$G = \int d^3x \, f_0(x) = \int d^3x \, \psi^\dagger \psi$$

which is the norm of the Dirac field. This also can be regarded as implying the conservations of probability for expectation values (cf. Greiner & Reinhardt 1996, p. 120).

These short examples for conservation laws in QFT have merely heuristic and no general value. However, it can be rigorously shown that the above results hold for QFT. In the case of the path integral formulation, Noether's theorem is then formulated in terms of the Ward-Takahashi identities (cf. Duncan 2012, p. 441 ff.). To be more precise, however, it has to be added that Noether's theorem is in general not valid in QFT. So-called anomalies lead to non-vanishing divergencies for the currents, that is, 5.3.49 will be non-zero. Nevertheless, since these anomalies arise from local gauge symmetry, they do not impugn the conserved quantities, energy in particular, that I will rely on in my further discussion. I will therefore not go deeper into this topic (cf. Duncan 2012, ch. 15.5).

I want to close this chapter by following Brading & Brown (2003) and emphasise the empirical significance of Noether's theorem. Spacetime symmetries can be interpreted actively, in that they tell us that if the laws of nature are valid in one system, they will also be valid in the transformed system. As Brading and Brown argue, however, this is not what gives Noether's theorem its empirical significance. The reason for this is that Noether's theorem is not restricted to spacetime symmetries, but also works with gauge symmetries, which cannot have the same active interpretation. Gauge transformations can in general not be connected to an empirical operation like going from one inertial frame into another. Instead, Brading and Brown make the point that the significance of the symmetries and hence Noether's theorem is comprised in what follows for the equations of motion.

> The imposition of a symmetry on a theory places a restriction on the possible form of the equations of motion of that theory, and insofar as this restriction has empirical significance then so too does the symmetry itself. *This* is the proper place to look when analysing the empirical significance of a given Noether symmetry. (Brading & Brown, 2003, p. 113)

On the one hand, we know from the validity of a symmetry that the equations of motion, that is the Euler-Lagrange equations, must transform according to these symmetries. On the other hand, the fields in the theory must be solutions to the Euler-Lagrange equations. So for example, if the equations of motion are symmetric under spacetime translations, then we know that the energy of the fields, solving the equations of motion, is conserved. This is where symmetries leave the realm of pure mathematics and gain empirical significance.[18]

5.3.4 The principle of locality

Locality, often called causality by physicists, is the implementation of special relativity into QFT.[19] It has the consequence that operators, evaluated at spacelike separated

[18] I want to emphasise that my only concern here is with Noether's theorem and that I am not making any claim about the empirical significance of symmetries in general. For a recent discussion of the question of the empirical significance of gauge symmetries see Greaves & Wallace (2011).

[19] For the obvious reason to avoid confusion, I will only use the term 'locality' to refer to the constraints of special relativity on QFT, and 'causality' for the philosophical concept describing

points in spacetime, commute and are therefore independent of one another. Thus, interactions in QFT happen locally, that is, the interaction evaluated at one point in space will have no effect on the interaction at other points that are spacelike separated. As such, locality is one of the essential properties of QFT, and with Haag (1996, p. 57) can be regarded as one of the postulates, which must stand at the beginning of a mathematical rigorous formulation of the theory.[20]

Locality is implemented in QFT as follows (ch. Greiner & Reinhardt 1996, ch. 4.4). In 5.3.27 already the equal time commutator relations for creation and annihilation operators have been given. These can be generalised from equal times to arbitrary times. Skipping the explicit calculation here, the commutator relations then are

$$\begin{aligned} \left[\hat{\phi}(x), \hat{\phi}(y)\right] =& 0 \\ \left[\hat{\phi}(x), \hat{\phi}^{\dagger}(y)\right] =& i\Delta(x-y) \end{aligned} \tag{5.3.51}$$

where $\Delta(x-y)$ is a Lorentz invariant function with the property of being zero for $(x-y)^2 < 0$, i.e., for spacelike distances.

This result is usually interpreted as saying that spacelike separated measurements can have no effect on each other, or in other words, that there cannot be any signalling with superluminal velocity. This interpretation, however, is problematic in two respects. First, it involves the notion of measurements, which, as is well known, has an unclear meaning and should be avoided. Second, as Weinberg (1995, p. 198) argues, even though some fields, like the electromagnetic field, can be measured at any of their points, this is not possible for other fields, like Dirac fields. For fields of the latter sort, it therefore cannot be said that spacelike separated measurements have no influence on each other, since there are no such measurements. Consequently, Weinberg opts for a more modest interpretation, and regards the commutator relations as only making the Lorentz invariance of the S-matrix explicit.

Even though I agree with Weinberg's critique on measurements, I want to argue that the commutator relations have a stronger consequence than merely the Lorentz invariance of the S-matrix. The interpretation favoured here must therefore not be regarded as standing in opposition to what Weinberg thinks, but merely as an extension. The first premise in the argument is that the S-matrix gives the probability for an interaction to happen. It then follows from locality that if an expression like, e.g., the vacuum expectation value (see 5.3.34) is evaluated for fields at spacelike separated points $x_1, \ldots, x_n$, then the S-matrix is zero and there will be no interaction. This constraint on any interaction is what the principle of locality implies in addition to Lorentz invariance. We therefore know more from locality than that the S-matrix is Lorentz invariant, but we also do not have to involve

causes, effects and their relation.

[20] It has to be added, though, that there might be processes that violate locality. However, these processes only occur in QFT in curved spacetime and algebraic formulations of QFT, neither of which I am concerned with in the present work, and have not been observed yet. I will therefore not study them here. The way that these processes violate locality are very subtle and discussed by Butterfield (2007).

measurements, since the significance of locality can be made clear by reference to interactions alone.

5.3.5 Virtual particles and force fields

The question of this section will be whether belief in the existence virtual particles is warranted or not. The debate over virtual particles[21] is dominated by arguments that can already be found in the debate between field and particle ontologies for QFT. In contrast to the latter debate, however, in the case of virtual particles it is not merely concluded that they are not particles in a literal sense, but that they do not exist at all. For example, Fox (2008, p. 49 f.) comes to the following conclusion:

> If virtual particles are of any use and meaning in quantum field theory, they are only instruments that by definition have to be unobservable and by nature unable to fly through our world on trajectories in space and time. Leaving aside any reasonable treatment of electrons, quarks, and other particles, virtual particles are a rather obvious case of entities not to be interpreted realistically at all.

As it stands this result can be interpreted as having the consequence that there is nothing between initial and final states in scattering events that mediates the force. In my opinion, this throws out the baby with the bathwater, and I will argue for the more balanced view that even though virtual particles are strictly not particles, there are intermediate states in quantum processes transmitting interactions. This section in roughly divided into two parts. I will first analyse virtual particles in the mathematical formalism of QFT and second, I will point out their role in experiments. The conclusion of this section will be that there is a sensible distinction between force fields, mediating the interaction, and virtual particles, and that only the former exist.

So where in the formalism of QFT do virtual particles appear? According to the Lagrangian 5.3.31 the interaction part of the Lagrangian for QED consists of the electromagnetic field $\hat{A}_\mu$, coupled to the current $j_\mu = \hat{\bar{\Psi}}\gamma^\mu\hat{\Psi}$. The contribution of the electromagnetic field to the S-matrix is given by the photon propagator

$$\langle 0|T\{\hat{A}_\mu(x_1)\dots\hat{A}_\nu(x_n)\}|0\rangle \ .$$

When evaluating the perturbation expansion of the photon propagator, one finds that integrals run over arbitrary momenta k_μ of the electromagnetic field, which therefore has photons off-mass shell, that is, photons for which $k_\mu k^\mu \neq 0$ (cf. Greiner & Reinhardt 1996, ch. 7.5). These off-shell photons are what is usually referred to as virtual particles.

The first point I want to make here is that belief in intermediate particles, understood in a very loose sense and not as small point-like objects, is warranted just as much as belief in other particles, which can occur in initial and final states. The

[21] See Fox (2008) for an overview.

reason is the group structure of QFT and the occurrence of the electromagnetic field in the Lagrangian. The main thrust of chapter 5.3.2 was that it is group structure telling us which kinds of different particles play a role in the theory. Accordingly, the electromagnetic field transforms under $U(1)$ symmetry and is thus massless with helicity ± 1. Since the electromagnetic field appears in the Lagrangian of QED, the corresponding particles, given by the group structure, are part of the processes described by the Lagrangian. In this respect, virtual particles are not different from any other particles.

However, in the discussion about virtual particles one usually goes a step further in that it is asked whether single parts of the perturbation series of the propagator can be regarded as describing real processes. The latter, as already mentioned, would describe processes with indefinite momenta corresponding to off-mass shell particles, which arguably is unphysical. (In this sense only, that is, for single parts of the perturbation series, I will henceforth use the term virtual particle. Again, the term 'particle' has to be understood in a very loose sense, since particles in quantum physics are very unlike their counterparts in classical physics.) Even more reasons why single parts of a perturbation series should not be interpreted on their own are given below.

Another aspect of the discussion about virtual particles is their operational significance, which is, e.g., emphasised by Falkenburg (2007, ch. 6.4.3). Highlighting the fact that all observations are the result of interactions and that virtual particles are an integral part of the theoretical description of interactions, Falkenburg concludes that in this respect virtual particles have an operational meaning. Even though Falkenburg's inference from experiments to the existence of virtual particles seems weak, she makes another important point here. Apart from special cases in which already the first order of a perturbation series gives precise predictions for experiments, it is only a superposition of the terms of the perturbation series that contributes to the scattering cross section. This leads Falkenburg to concede that single virtual particles have no operational meaning, but only all the virtual particles of a perturbation series together:

> Virtual particles are formal tools of the perturbation expansion of quantum field theory. They do not exist on their own. Nevertheless, they are not fictitious but rather produce *collective effects* which can be calculated and measured with high precision. (Falkenburg 2007, p. 238)

There are some objections against Falkenburg's reasoning. The most common argument against the existence of virtual particles is the argument from superposition (cf. Weingard 1982, 1988; Teller 1995). This argument simply points out that the perturbation series does not contribute as a sum of its terms to the scattering cross section, but only as a superposition. More precisely, we are not talking about a superposition of properties, as for example one particle can be in a superposition of different spin states, but what is superposed are multiple objects. It is then noted that it is unclear what a superposition of particles is, which has the conclusion that we cannot allow such a superposition in our ontology. This problem, however, can be answered in the same way as Falkenburg does, namely in acknowledging that only the collection of virtual particles is real, thus evading the picture that a superposition is somehow created out of single particles, each existing in their own right (cf. Teller 1986). Superpositions are therefore more a problem for a literal particle interpretation, since it is indeed hard to imagine how particles taken in a literal sense can be superposed.

However, one might insist that it is not entirely clear how Falkenburg can speak of a collection of something, if what the collection is composed from has no existence. Even though Falkenburg does not say this explicitly, the answer seems to be that 'collection' does not refer to an aggregation of localised objects, since virtual particles are "*neither local nor localizable* by a particle detector" (Falkenburg 2007, p. 236), but rather to virtual particles in so far they are "*discontinuous*, i.e., come as quanta" (ibid.). The upshot from Falkenburg is therefore, that as a result of their experimental significance virtual particles have to be real, but this does not amount to them being local objects. Consequently, a superposition of virtual particles is not a superposition of local objects, but might be better understood as a superposition of properties of one object.

More serious problems arise from the following considerations, though again I will argue that these problems only speak against virtual particles, but not against intermediate states. Since virtual particles are off-mass shell they can have arbitrary high energies, which is clearly unphysical. What is more, the infamous divergencies of QFT and thus the need for renormalisation arise exactly in the calculation of virtual particle terms. Consequently, all the problems or possible interpretations that come with renormalisation are also problems for virtual particles (cf. Cao 1997, p. 204). The common core of these two problems, however, is that virtual particles only enter the formalism when perturbative methods are used. The obvious question then is, whether virtual particles, taken separately or as a superposition, are only artefacts of perturbation theory. Two points speak in favour of an affirmative answer. The first is that in the definition of the propagator, e.g. in 5.3.35 or 5.3.43, virtual particles do not appear. It is only because this expression cannot be calculated analytically that perturbative methods have to be used. Second, there are non-perturbative formulations of QFT, such as the path integral formalism, in which the question of virtual particles consequently makes no sense (cf. Maggiore 2005, p. 219). As Rohrlich (1999, p. 363) sums it up:

> Virtual particles arise in perturbation expansion, i.e. in a particular approximation method. Other methods of solving the problem (for example of finding the scattering amplitude) do not necessarily involve them.[22]

As a consequence, not only single virtual particles do not exist, but also their superposition. Therefore I cannot agree with Falkenburg in emphasising the collective effects of virtual particles. The effects can neither be due to virtual particles nor to a collection of them.

[22]See also Redhead (1982) for the same point.

However, it is important to emphasise that this conclusion does not lead to discarding the force fields as the transmitter of interactions as a whole. Force fields, like the Maxwell field in QED, are an indispensable part and contribute to every observable effect of every formulation of Lagrangian QFT. We can thus banish virtual particles completely, while at the same time agreeing with Zee (2003, p. 27) in saying that "the exchange of a particle can produce a force was one of the most profound conceptual advances in physics." Again, if 'particle' here is understood merely in a metaphorical sense, then it is unproblematic to agree with Zee that the description of interaction via forces was indeed a great advance in physics. I will from now on refer to these force fields as intermediate fields or states.[23]

5.3.6 Feynman diagrams

It is hard to imagine today's practice in physics without Feynman diagrams. They are used everywhere as a tool to handle complicated algebraic calculations. What is more, their ubiquitous presence as well as their pictorial character has lead some to believe that they are more than mere tools and indeed show how nuclear processes look like. This section will discuss Feynman diagrams between these two poles: tool and pictorial representation. I will first present how the diagrams enter physics and then go through the arguments that have been put forward for one or the other of the two possibilities. The conclusion will be that Feynman diagrams cannot be regarded as giving insights into nature in their own right, beyond what is already known from the formalism of QFT without diagrams.

The place where Feynman diagrams appear in the canonical quantisation formalism is the perturbation expansion of the S-matrix, 5.3.33, which consists of time-ordered products of field operators.[24] According to Wick's theorem (cf. Greiner & Reinhardt 1996, ch. 8.5) such products can be reformulated and computed through sums of contractions, defined as

[23] A topic that I don't want to go in further here is the very popular image of the vacuum filled with virtual particles constantly coming into existence and vanishing again. Furthermore, it is often supposed that they have a measurable effect, namely the Casimir effect, by putting pressure on two plates close to each other. I have only brief comments on that. First, contrary what seems to be the popular understanding, the vacuum state is not literally nothing, but merely the ground state of a field. Second, fluctuations of the vacuum state are due to the Heisenberg uncertainty principle and thus a property of every state in quantum physics. Virtual particles coming into existence and vanishing after some time again would have to be described by an appropriate dynamics, that is, a state and an evolution operator for that state. The latter, however, are not part of contemporary QFT. Thus, vacuum fluctuations are fundamentally different from, e.g., particle creation and annihilation in scattering experiments. Third, it is not at all clear whether virtual particles, or even vacuum fluctuations are part of the explanation of the Casimir effect (cf. Jaffe 2005).

[24] Here I only intend to describe Feynman diagrams as they were used originally. Today they are applied in many different places, which I will not present, since my argument does not depend on it. Kaiser (2005) gives a detailed discussion of the migration of Feynman diagrams from their origins into other areas, and captures this under the heading of the 'plasticity' of the diagrams.

$$\hat{\phi}_1(x_1)\hat{\phi}_2(x_2) := \langle 0|T\left\{\hat{\phi}_1(x_1)\dots\hat{\phi}_2(x_2)\right\}|0\rangle$$

i.e., vacuum expectation values.

Taking again QED, with the Lagrangian 5.3.28, as an example, the S-matrix can be expanded as (cf. Greiner & Reinhardt 1996, ch. 8.6)

$$\begin{aligned}
S =&\mathbb{I} + \sum_{n=1}^{\infty} S^{(n)} \\
=&\sum_{n=1}^{\infty} \frac{1}{n!}(-ie)^n \int d^4x_1 \dots d^4x_n\, T\left[: \hat{\bar{\Psi}}(x_1)\gamma_1^{\mu}\hat{\Psi}(x_1)\hat{A}_{\mu_1} \cdots \hat{\bar{\Psi}}(x_n)\gamma_n^{\mu}\hat{\Psi}(x_n)\hat{A}_{\mu_n} :\right]
\end{aligned}$$

with $: \dots :$ denoting the normal product in which negative frequency parts of the operators are always written on the left of positive frequency parts. Now one finds that the first order term does not contribute to the S-matrix. The second order term, however, does contribute and is explicitly given as

$$\begin{aligned}
S =&\frac{1}{2!}(-ie)^2 \int d^4x_1 d^4x_2 \; : \hat{\bar{\Psi}}(x_1)\gamma^{\mu}\hat{\Psi}(x_1)\hat{\bar{\Psi}}(x_2)\gamma^{\nu}\hat{\Psi}(x_2)\hat{A}_{\mu}(x_1)\hat{A}_{\nu}(x_2) : && \text{(a)} \\
&+\frac{1}{2!}(-ie)^2 \int d^4x_1 d^4x_2 \; : \hat{\bar{\Psi}}(x_1)\gamma^{\mu}\hat{\Psi}(x_1)\hat{\bar{\Psi}}(x_2)\gamma^{\nu}\hat{\Psi}(x_2)\hat{A}_{\mu}(x_1)\hat{A}_{\nu}(x_2) : && \text{(b)} \\
&+\frac{1}{2!}(-ie)^2 \int d^4x_1 d^4x_2 \; : \hat{\bar{\Psi}}(x_1)\gamma^{\mu}\hat{\Psi}(x_1)\hat{\bar{\Psi}}(x_2)\gamma^{\nu}\hat{\Psi}(x_2)\hat{A}_{\mu}(x_1)\hat{A}_{\nu}(x_2) : && \text{(c)} \\
&+\frac{1}{2!}(-ie)^2 \int d^4x_1 d^4x_2 \; : \hat{\bar{\Psi}}(x_1)\gamma^{\mu}\hat{\Psi}(x_1)\hat{\bar{\Psi}}(x_2)\gamma^{\nu}\hat{\Psi}(x_2)\hat{A}_{\mu}(x_1)\hat{A}_{\nu}(x_2) : && \text{(d)} \\
&+\frac{1}{2!}(-ie)^2 \int d^4x_1 d^4x_2 \; : \hat{\bar{\Psi}}(x_1)\gamma^{\mu}\hat{\Psi}(x_1)\hat{\bar{\Psi}}(x_2)\gamma^{\nu}\hat{\Psi}(x_2)\hat{A}_{\mu}(x_1)\hat{A}_{\nu}(x_2) : && \text{(e)} \\
&+\frac{1}{2!}(-ie)^2 \int d^4x_1 d^4x_2 \; : \hat{\bar{\Psi}}(x_1)\gamma^{\mu}\hat{\Psi}(x_1)\hat{\bar{\Psi}}(x_2)\gamma^{\nu}\hat{\Psi}(x_2)\hat{A}_{\mu}(x_1)\hat{A}_{\nu}(x_2) : && \text{(f)} \\
&+\frac{1}{2!}(-ie)^2 \int d^4x_1 d^4x_2 \; : \hat{\bar{\Psi}}(x_1)\gamma^{\mu}\hat{\Psi}(x_1)\hat{\bar{\Psi}}(x_2)\gamma^{\nu}\hat{\Psi}(x_2)\hat{A}_{\mu}(x_1)\hat{A}_{\nu}(x_2) : && \text{(g)} \\
&+\frac{1}{2!}(-ie)^2 \int d^4x_1 d^4x_2 \; : \hat{\bar{\Psi}}(x_1)\gamma^{\mu}\hat{\Psi}(x_1)\hat{\bar{\Psi}}(x_2)\gamma^{\nu}\hat{\Psi}(x_2)\hat{A}_{\mu}(x_1)\hat{A}_{\nu}(x_2) : && \text{(h)}
\end{aligned}$$

To obtain a result from this long expression, it is now convenient to first draw Feynman diagrams according to the following Feynman rules:

1. Every point x corresponds to a vertex, that is, a point where other lines meet.
2. Each fermion field operator corresponds to an external straight line, directed either into a vertex or coming from one.
3. Each boson field operator corresponds to a external wavy line, without direction, but with a vertex on one end.

4. Each contraction between two fermion field operators, $\underbrace{\hat{\Psi}(x_1)\hat{\bar{\Psi}}}(x_2)$, corresponds to an internal line with direction from x_1 to x_2.

5. Each contraction between two boson field operators, $\underbrace{\hat{A}_\mu(x_1)\hat{A}_\nu}(x_2)$, corresponds to an undirected[25] internal wavy line between x_1 and x_2.

Accordingly, term (a) above can be translated into two diagrams, each with two external fermion and one external boson line. For all eight terms of the expansion above, this leads to the following schematic diagrams, where time is the vertical axis:

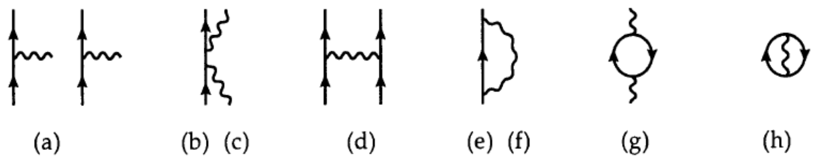

Figure 5.2: Schematic Feynman diagrams, corresponding to the second order perturbation expansion of QED. In: Greiner & Reinhardt (1996, p. 239)

The specific form of the contributing diagrams depends then on the interaction one is describing. If, for example, the process one wishes to compute is the scattering between two electrons or positrons or one electron and one positron, then only diagram (d) will contribute to the S-matrix, since only this diagram has the right number of particles in the initial and final state. More precisely, depending on whether one considers either (i) electron-electron scattering, or (ii) positron-positron scattering, or (iii) electron-positron scattering, the schematic diagram (d) above will take on one one of the following forms:

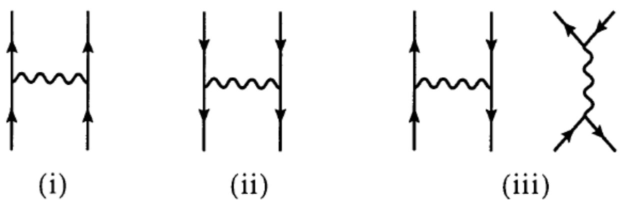

Figure 5.3: Feynman diagrams corresponding to term (d) of the second order perturbation expansion of QED. In: Greiner & Reinhardt (1996, p. 240)

Finally, each element of a diagram corresponds to an algebraic expression, e.g., internal wavy lines to the Feynman propagator for photons, which can then be set together to give the entire S-matrix.

Since the appearance of Feynman diagrams in the path integral formalism is very similar to that in the canonical quantisation formalism, I will not go into details here. Suffice it to note that from the perturbative expansion of the Green's function

[25] This is due to the symmetry of the photon propagator, that is, $D_{\mu\nu}(x_1 - x_2) = D_{\mu\nu}(x_2 - x_1)$.

5.3.42 Feynman diagrams can be created according to the following rules: (i) The number of vertices is given by the order of the perturbation series. (ii) Propagators correspond to internal lines. (iii) Source term correspond to external lines. Of course, since we are now in the path integral formalism, propagators are now Feynman path integrals of classical field configurations, and source terms also consist of classical fields only.

I will now turn to the question how Feynman diagrams have to be interpreted, that is, in an instrumentalist or realist fashion. The first argument I want to consider is directed against the realistic interpretation of Feynman diagrams in general. Similar to virtual particles, Feynman diagrams only show up in the perturbative approach to QFT. Since perturbative methods are not compulsory and can often be avoided, one might want to conclude that the diagrams cannot contribute to the interpretation of QFT. This conclusion, however, would be overly hasty. The perturbation expansion is not completely detached from the rest of the formalism, that is, some parts of the structure of QFT are also visible in the expansion. One therefore cannot exclude the possibility to learn something about QFT in general from the perturbative method. Likewise, in the preceding section I have only argued against a literal understanding of virtual particles as particles, but not against their existence as forces.

Consequently, I will follow Wüthrich (2012) and adopt a more cautious approach in answering the question on Feynman diagrams. As he notes, there is not only the choice between either dismissing or affirming Feynman diagrams as a whole, but there is also the possibility that they partially represent the world.[26] More precisely, diagrams have at least three properties, each of which may independently of the others either represent or not. The properties are: 1) The diagrams have parts, that is, they consist of external lines, internal lines and vertices. 2) The parts are connected at the vertices. 3) The parts are localised in space. In what follows I will introduce the arguments for different interpretations of the diagrams based on this distinction.

The first trait of Feynman diagrams I want to discuss is their localisation in spacetime, i.e., the property of being constructed out of lines and points. Harré (1988) highlights that Feynman diagrams resemble particle tracks as they show up in measurement devices like cloud chambers and this, as Harré argues, might be the most important reason why one might mistakenly believe that the diagrams give an exact picture of physical processes. That this belief indeed would be a mistake has several reasons. First, as I have argued in chapter 5.2.3, tracks in cloud chambers are not particle tracks at all, but rather a series of interactions. What is more, the particles whose tracks supposedly compose the diagrams have a well defined momentum and thus, according to the Heisenberg uncertainty principle cannot have a well defined position at the same time (cf. Meynell 2008). Thus, it follows at least that particles cannot be as exactly localised as the lines in the diagrams.

[26]At this point, I am not interested in a detailed account of how partial representation could work. Suffice it to say that there are worked out theories of partial representation, for example as given by da Costa & French (2003).

Next is the property of the diagrams having parts, which is independent of the fact that these parts cannot be as localised as well as in the diagrams. And indeed, this is a property in which Feynman diagrams represent processes, since it rests on the structure of the Lagrangian. The straight and wavy lines in the diagrams correspond to the fermion and boson fields, appearing in the Lagrangian. The distinction between external and internal parts, on the other hand, is determined by which process is calculated and consequently which particles appear in the initial and final states and which one only in the intermediate state.

Also the property that the parts are connected at vertices can at least partially be seen to represent. The first Feynman rule, stated above, tells that the number of vertices equals the number of space-time coordinates in the perturbation term. This can be regarded as reflecting the locality of the Lagrangian, that is, interactions between fields are always evaluated point by point at one space-time coordinate only and there is no action at a distance. However, since the space-time coordinates are integration variables and usually the integration runs over all of space-time, the vertices cannot represent insofar as they are points. Rather, they only represent locality and nothing more.

Wüthrich (2012, p. 178), however, goes a step further and claims that Feynman diagrams “allow us to explain certain aspects of the behaviour of the system which they represent”. According to him, diagrams like that on the right-hand side of picture (iii) in figure 5.3 for electron-muon scattering “allow us to explain that, most of the time, the electron and the muon deflect each other only a little.” (Wüthrich 2012, p. 178) This explanation supposedly rests on the facts that long range forces lead to a high reaction rate and that a force can only have a long range if the electron and muon deflect each other only little. Nevertheless, I have to disagree with Wüthrich when he claims that it is the Feynman diagram that explains the specific deflection angle. The problem is that the premises and the conclusion of Wüthrich’s argument are not part of the Feynman rules and are thus not used in their construction of diagrams nor can they be read of from them. The range of the forces, the angles between meeting lines and so on have no importance for the construction of Feynman diagrams and thus the diagrams cannot represent those features of the world. Even though Wüthrich’s explanation is correct, the explanation can be given perfectly without any reference to Feynman diagrams. Therefore, it is not the Feynman diagram doing the explanatory work here and they do not represent that what is explained.

In conclusion, Feynman diagrams can represent certain features of physical processes, however, they are of little help for the interpretation of QFT. Since the properties of the diagrams that do represent are all derived from the underlying formalism, diagrams simply tell us nothing new about the physical process. Instead of taking the detour of Feynman diagrams, any interpretation of QFT should rely directly on the mathematical formalism. I will therefore not consider them further in this work.

5.3.7 Probabilities

This chapter aims to clarify the role of probabilities in QFT. The most basic routes that one can take in explaining the role of probabilities in scientific theories are summarised by Wüthrich (2011, p. 365):

> There are at least two roles that probabilities can play in a dynamical theory. First, they may codify a distribution over initial or boundary conditions. Second, probabilities may concern the dynamical evolution given a certain initial state of the physical system under consideration. Orthogonally, probabilities in either role may be objectively in the world, i.e. real-world chances, or they may be subjective and arise only due to our ignorance of the exact state of affairs in the world.

Following these distinctions, I will argue first that QFT is a probabilistic theory, because it describes the probabilistic dynamical evolution of a system and second that these probabilities are objective. To this aim, I will first give a very brief overview of the probability calculus, which is the mathematical fundament for discussion of probabilities. I will then show how QFT satisfies this calculus and finally give arguments for the objectivist interpretation. I will not however, go further in the interpretive realm, that is, I will not discuss the two mainstream views on objective probabilities, namely frequentism and propensities. Thus, I am content with showing that QFT is an objective probabilistic theory, without giving any further insight about where these probabilities come from.[27]

Based on Kolmogorov (1956, § 1), Gillies (2000, p. 59) defines probabilities as follows:

> Let E, F, ..., E_1, ... stand for events, concerning which we can have some degree of belief whether they will occur, or have occurred. Let Ω denote the certain event, which must occur. There are then three axioms of probability.
>
> 1. $0 \leq P(E) \leq 1$ for any E, and $P(\Omega) = 1$.
> 2. (Addition Law) If E, ..., E_1 are events which are exclusive (i.e. no two can both occur) and exhaustive (i.e. at least one must occur), then $P(E_1) + \ldots + P(E_n) = 1$
> 3. (Multiplication Law) For any two events E, F $P(E \,\&\, F) = P(E \mid F)\, P(F)$

For present purposes conditional probabilities are of no importance, and thus the third axiom can be omitted. I will now show how QFT fulfils the first two axioms.

QFT is a theory to describe scattering events, that is, the evolution of an initial to a final state through interactions (and the very similar decays). The basis of scattering theory is that we have two different dynamics for the same system, the interacting dynamics $\hat{H}^I$ and the free dynamics $\hat{H}^0$. If Σ is the set of all states (rays

[27] Furthermore, I am not arguing that *all* probabilities that appear in quantum physics are Kolmogorovian in the sense explicated below. Rather, there are places, at least in QM, where non-classical probabilities appear (cf. Rédei & Summers 2007).

in Hilbert space), then there are pairs of states $< \psi_i, \psi > \in \Sigma$ and $< \psi_f, \psi > \in \Sigma$ so that $\lim_{t\to-\infty}(\hat{H}^I\psi - \hat{H}^0_{in}\psi_i) = 0$ and $\lim_{t\to+\infty}(\hat{H}^I\psi - \hat{H}^0_{out}\psi_f) = 0$. For every ψ_i there has to be only one ψ and vice versa. In other words, a state ψ should transform by the interacting Hamiltonian to ψ_i for the past and to ψ_f for the future. In general the initial and the final state will have different dynamics, that is, $\hat{H}^0_{in} \neq \hat{H}^0_{out}$. We can then define two subsets of Σ

$$\Sigma_{in} = \left\{ \psi \in \Sigma | \exists \psi_i \in \Sigma \text{ with} \lim_{t\to-\infty} \hat{H}^0_{in}\psi_i - \hat{H}^I\psi = 0 \right\}$$

$$\Sigma_{out} = \left\{ \psi \in \Sigma | \exists \psi_f \in \Sigma \text{ with} \lim_{t\to+\infty} \hat{H}^0_{out}\psi_f - \hat{H}^I\psi = 0 \right\} .$$

We can then go on and define the S-operator as a linear unitary transformation from Σ_{in} onto Σ_{out}. In QFT this is the time evolution operator $\lim_{t_2\to+\infty}\lim_{t_1\to-\infty}\hat{U}(t_2,t_1)$ (cf. Reed & Simon 1979, ch. XI, sc. 1).

Now, the scattering matrix S "describes the probability amplitude for a process in which the state makes a transition from an initial state to a final state under the influence of an interaction." (Greiner & Reinhardt 1996, p. 219) That is:

$$S_f = \langle\psi_f|\hat{S}|\psi_i\rangle = \lim_{t_2\to+\infty}\lim_{t_1\to-\infty}\langle\psi_f|\hat{U}(t_2,t_1)|\psi_i\rangle$$

With respect to the axioms of the probability calculus, the amplitude S_f is an event, a complex number, the probability of which is given by its absolute square

$$P(S_f) = |\,\langle\psi_f|\hat{S}|\psi_i\rangle\,|^2 \quad , \quad P(S_f) \in \mathbb{R}|0 \leq P(S_f) \leq 1 \, .$$

Thus the first axiom is fulfilled. If we now sum up over all possible outcomes ψ_f, the probability is always one

$$\sum_f P(S_f) = \sum_f |\,\langle\psi_f|\hat{S}|\psi_i\rangle\,|^2 = 1 \, ,$$

which fulfils the second axiom. It therefore follows that QFT is a probabilistic theory in the sense that its laws, not the initial or boundary conditions, are probabilistic. The next question is, how these probabilities have to be interpreted, either objectively or subjectively.

The usual argument for objective probabilities in physics is that physics, at least for the realist, gives an objective description of the world. Therefore, if the laws of physics are probabilistic, then probabilities are part of the world, and not a mere artefact of our knowledge of the world (cf. Krips 1989). As Earman (1986, p. 148) puts it:

> I will simply ignore the contentions of those probability theorists who assert that all probability statements ultimately refer to subjective or personal degrees of belief. The theories of physics purport to give us truths about probabilities that are objective and observer independent, and it is the nature of these probabilities we seek to understand.

Before analysing how the probabilities in QFT fare in the light of this argument, it will be helpful to have a quick look at the situation in QM. As will become clear, under a certain interpretation of QM the interpretation of probabilities in this theory differs significantly from that in QFT. Giving a brief presentation of QM thus may serve as a contrast to QFT sharpening the understanding for the latter. At the same time, however, I do not intend to do more than that and do not endorse any view about QM here.

Probabilities in QM appear in the Born rule, and lead, e.g., Auyang (1995, p. 85) to conclude that QM "is essentially probabilistic." However, as is well known, the Born rule is incompatible with a universally valid Schrödinger equation. What is more, the Born rule is usually interpreted as giving the probability for certain measurement outcomes, which suggests that it is only a rule about what we can know about nature, rather than how nature really is.[28] Thus, the Born rule does not seem to be part of the physical theory of the world and the above argument for objective probability does not apply to it. It follows that under the above interpretation of the Born rule, probabilities are not part of world.[29] As a side note, another example, analogous to the status of the Born rule in QM, is that of thermodynamics. Here again the laws of the theory are probabilistic, however, since thermodynamics is not a theory about fundamental physical properties, the probabilities are prima facie merely subjective. In short, whenever a theory or parts of a theory are not about fundamental physical facts, the above argument for objective probability does not apply, and the probabilities, occurring in these theories are more likely to be subjective.

Are the probabilities in QFT similar to the Born probabilities, or can a different interpretation be given? The first problem of the Born rule was that it stands in contradiction to the Schrödinger equation. While the Schrödinger equation gives a unitary evolution of the wave-function for all times, the Born rule implies a collapse of the latter, i.e., a non-unitary process. In QFT, however, the situation is different. Here the S-matrix does not contradict the free dynamics of the fields, since when interacting, the fields are evolving according to the interaction dynamics.[30] The S-matrix and therefore the probabilities in QFT do not stand in opposition to the rest of the theory, but rather are the central ingredient of QFT. Now, the second problem of the Born rule, as applying only to measurements, was that its interpretation is unclear, and again I want to urge that this problem does not arise in QFT. The latter is the description of scattering events and decays, i.e., is the evolution of an initial into a final state, and the S-matrix gives the probability for these events to happen. Transition probabilities therefore have a clear physical meaning and are

[28]See section 6.3 below for a detailed discussion of the Born rule and the measurement problem.

[29]This short passage is of course not a complete discussion of the problem whether QM is probabilistic. Another relevant question in QM is, e.g., whether the Schrödinger equation is indeed deterministic (cf. Wüthrich, 2011 for a discussion), and Krips (1989) argues that the probabilistic nature of QM can be inferred from interference effects.

[30]This is not to say that the relation between free and interaction dynamics is completely unproblematic. I will discuss these matters in section 6.2 below.

free from the ambiguities the term 'measurement' is beset with.[31]

The upshot of this discussion is that Earman's argument for objective probabilities, mentioned above, applies to QFT. First, according to a realistic interpretation of QFT, this theory provides an objective, observer independent description of the world. Second, probabilities are an ubiquitous part of the laws of that theory. Therefore, the probabilities that occur in QFT are objective probabilities, or in other words, QFT is objectively probabilistic.

5.4 Conclusion

The motivation of this chapter was to analyse whether there are features of QFT that look promising as a reduction basis for causation. The guiding idea in the background was Dowe's CQT, and if possible, it would have been the easiest way to simply leave the CQT as it is. However, Dowe defines causation as the exchange of a conserved quantity at spacetime points where world lines meet, and the relevant world lines were singled out by the ontology of the theory. None of the concepts that are part of this definition can be retained in the light of QFT. Instead, summarising the results of the previous sections, I want to put forward the following properties as the main building blocks for causation in QFT. In the chapter 7 I will then put these results at work and form with them a new version of the CQT.

The first problem was in how far the notion of world lines and the meeting of world lines could be made sense of in QFT. Here a preliminary result was that even in field theories, with fields defined over all of spacetime, world lines that do not fill out the whole of a Minkowski diagram can be defined by only taking a region of spacetime into account where the field has a certain value. This proposal then was concretised following the work of Haag, Swieca and Wallace as regarding a quantum field to be localised in a region where it significantly differs from the vacuum state. This definition of localised fields also agrees with experimental practice. In scattering experiments fields are localised by the use of collimators in a very small region and tracks in cloud chambers, even though they are not world lines, but a series of interactions, show that what is interacting has to be localised in a small region as well. However, even though localised fields have their applications in experiments, their general significance remained somewhat unclear. There is certainly no constraint in QFT for fields to be always localised in a small region. On the other hand, the principle of locality proved one of the most basic constraints in QFT. It therefore looks possible to replace Dowe's intersecting world lines by the principle of locality. Just as the meeting or world lines in a Minkowski-diagram was supposed by Dowe, the principle of locality teaches us that interaction goes hand in hand with some sort of intersection and is never action at a distance.

Next, I have presented the role of group theory in QFT. The result was that

[31] More details on why exactly the Born-rule has an ambiguous interpretation and why this is not the case for QFT will be given in the discussion of the measurement problem in section 6.3 below.

whatever takes part in an interaction in QFT has to be an irreducible representation of one of the relevant symmetry groups, singled out by the specific values of rest-mass and spin, or consist of irreducible representations, if the interacting object is not an elementary particle. This suggests that, contrary to Dowe, we can refer to the relata of causation without invoking a complete ontology for QFT, which arguably does not exist at the moment. While Dowe simply pointed towards the objects in the ontology of science as the relata of causal relations, but gave no further explanation on what they are, we can hopefully now be more specific.

Furthermore, the theoretical description of interaction in QFT has shown that interactions are not simple intersections of whatever interacts, but are always mediated by a force. Even though the conclusion about the existence of virtual particles was negative, this does not imply that force fields in general cannot play the role of intermediate fields in interactions. I will argue later, that this sheds some light on what the relation between cause and effect is. Furthermore, we found that energy is locally conserved in all interactions in QFT, which strongly suggests that it is the only property that a transference theory of causation can be based on. Finally, I have argued that QFT is a probabilistic theory, from which follows that the account of causation developed later has to be probabilistic.

Overall, QFT seems to have very promising characteristics for developing a causal interpretation. Even though it is clear that Dowe's CQT cannot be saved, the central idea that causation is the transfer of a conserved quantity might be carried over into a new theory of causation. This will be the topic in chapter 7. But before that, I will discuss what I believe are the three main objection to a causal interpretation of QFT.

6 Physics has bad news (or so it seems)

6.1 Introduction

The picture of interactions in QFT that emerges from the last chapter shows that interactions can be regarded as processes whereby energy is exchanged within a well localised area of spacetime. I am aware that there are many good objections to this presentation of QFT and consequently to the conclusions I ended up with. I cannot possibly discuss all these objections to the extent that they deserve. For example, there is an important dispute about which formulation of QFT, axiomatic or textbook, should serve as the basis of an interpretation, which I will only briefly have a look on in section 6.2.2. Another example of a convolute of possible problems, which I will not touch at all, is the relation between QFT and a theory of gravity.[1] Yet another serious problem might be the supposed non locality of the Aharonov-Bohm effect.[2] That said, in this chapter I will discuss three problems, of which I think they are most directly threatening the interpretation of QFT that I defend. These problems are largely distinct from each other, and as a result this chapter consists of three independent sections.

The first problem is Haag's theorem, according to which it is mathematically impossible to describe interactions between fields. If this was true, then QFT could hardly be interpreted as describing causal interactions between fields. The second problem is the infamous measurement problem. Arguably, it can only be solved by moving away from standard QM, and on towards one of a range of alternative theories that supposedly have very different philosophical interpretations. Finally, the last problem is that of whether entanglement is a causal relation. If the answer to that question is affirmative, then the prospects for a transference theory of causation are extremely dim, given that no transfer takes place between the measurements on an entangled state. The discussion of entanglement will be somewhat removed from QFT and only take place on the level of QM. This follows the standard approach in the literature, as the importance of entanglement in QFT is still unclear to some extent, as can be seen from the discussions around the Reeh-Schlieder theorem (cf. Huggett 2000, sec. 4.3). Nonetheless, I think it is reasonable to assume that

[1] I wish to highlight that, in a very similar approach to the one followed here, Bartels & Wohlfarth (2014) argue that in General Relativity causation can be reduced to the asymmetric flow of energy.

[2] See Wallace (2014) who argues that, contrary to the usual understanding, the Aharonov-Bohm effect does not show any new physics over standard quantum electrodynamics.

entanglement as a well established experimental fact will have some QFT analogue and therefore needs to be discussed here.

6.2 Haag's theorem

6.2.1 What it is and why it is worrying

In my exposition of QFT I happily made use of the interaction picture. This approach roughly consists of the assumptions that the complete Hamiltonian of QFT can be separated into a free and an interaction part, $\hat{H} = \hat{H}_0 + \hat{H}_{int.}$ (cf. equation 5.3.32), and that the fields in their initial states are free, described by $\hat{H}_0$, then interact according to the full Hamiltonian, $\hat{H}$, and then evolve into their free final states again. This is a standard method in QFT, applied by most if not all introductory QFT textbooks. As is well known, however, the interaction picture is not as innocuous as it might appear given its prevalence. In this section and the next, I wish to explain what the problem exactly is, and how it can be circumvented.

The proof, showing that something is wrong with the interaction picture, is now known as Haag's theorem.[3] The thrust of it is that free and interacting fields cannot be described in the same Hilbert space and that there is no unitary transformation from one Hilbert space to the other; free and interacting fields are unitarily inequivalent. Under the usual interpretation of unitary equivalence, that is, "unitary equivalence [...] guarantees physical equivalence" (Earman & Fraser 2006, p. 306), the importance of this result is immediately clear. There is no physical process that would turn free fields into interacting fields or vice versa. Given that for this reason the interaction picture is mathematically not sound, according to Fraser (2008, p. 847) it can be concluded "that if all the assumptions of the interaction picture are accepted, then [...] the fields that were intended to describe the interaction — actually describe a free system."

With respect to the present work, this result generates two different problems, one for the interpretation of QFT in general, and one for the causal interpretation in particular. As Duncan (2012, p. 359) summarises it, the former problem stems from "the understandable fear that perturbative methods, and the Feynman graph approach [...], are founded on mathematical quicksand, and therefore in some sense unreliable despite their obvious empirical success over the years." This, of course, is just a part of the broader topic whether textbook QFT, even though not always mathematically rigorous, is in some sense still a 'good' theory, or whether some axiomatic approach to QFT has to be preferred. I will have something to say on this debate in the next section, though, since it is not directly in my focus, I can hardly do more than to say which side I stand on and why.

The second problem comes from the implications Haag's theorem has for the causal interpretation of QFT. Looking (maybe) naively on the interaction picture,

[3]See Earman & Fraser (2006) for a demonstration of the proof and the history of Haag's theorem, pointing out that it should not be credited to Haag alone.

one could assume that it describes the continuous time evolution of quantum fields from their initial free states through the interacting states to the final free states, and thus the complete time evolution of a causal interaction. The basis of this interpretation can most clearly be found in equation 5.3.33, that is the S-matrix element from which the scattering cross section can be calculated, where the operator $\hat{S}$ is simply the unitary time evolution operator. From Haag's theorem, however, we know now that there is no continuous time evolution that would evolve a free field into an interacting field. Or in other words, in so far QFT describes the time evolution of fields that are in an initial free state, the fields have to be free at all times. I assume it to be rather obvious that one would have difficulties to interpret free fields as causally interacting. In particular for transference theories of causation this is a problem, because free fields are not transferring any conserved quantity to other fields. Nevertheless, I believe that both problems can be answered, giving support to textbook QFT and the causal interpretation, and will explicate why in the next section.

6.2.2 How to live with it

Given the nature of Haag's theorem as a mathematical proof, one can hardly expect to show that is it wrong, but maybe we can live with it nonetheless. According to Earman & Fraser (2006, p. 306), "Haag's theorem does not show that no quantum field theory exists which differs from a free field theory". Thus, it might be the case that its implications are more restricted than initially thought. This, in general terms, will be the strategy applied here of how to cope with Haag's theorem.

As both Earman & Fraser (2006) and Duncan (2012) point out, there are two well-known methods for calculating quantum field theoretical interactions, to which Haag's theorem does not apply. The first method is the usual perturbative evaluation of interactions. As it turns out, the introduction of high and low energy cutoffs in the perturbation series make the application of Haag's theorem impossible and allow for the calculation of the S-matrix in the interaction picture. This is understandable, given that Haag's theorem is a consequence of quantising over an infinite spatial box. Thus, "the same regularizations needed to make proper mathematical sense of the dynamics of an interacting field theory at each stage of a perturbative calculation will do double duty in restoring the applicability of the interaction picture at intermediate stages of the calculation." (Duncan 2012, p. 370)

The second reply to Haag's theorem is to point out that "Haag–Ruelle scattering theory is available as a mathematically consistent alternative framework for scattering theory." (Earman & Fraser 2006, p. 333) As Duncan (2012, p. 363 f.) sums it up, "there is no difficulty whatsoever in establishing a well-defined unitary relation between the in- and out-states of an interacting field theory. [...] Haag-Ruelle and LSZ scattering theories lead to a perfectly well-defined, and *unitary*, S-matrix, on the basis of exactly the same axiomatic framework which can be used to establish the validity of Haag's theorem." Contrary to the method used in the interaction picture, in Haag-Ruelle scattering theory the free fields are constructed from the interacting fields as their asymptotic states, and both, the free and interacting fields, can be described in the same Hilbert space. However, the free field of Haag-Ruelle scattering theory is not the free field of the interaction picture, which is governed only by $\hat{H}_0$. In particular, since in Haag-Ruelle scattering theory the self interaction is never completely switched off, the two fields will have a different mass and different vacuum expectation values (cf. Reed & Simon 1979, p. 317; Ticciati 1999, ch. 10.4). Nevertheless, they easily fit into the intuitive description of free fields as it is given in the interaction picture, since for example initial and final states in Haag-Ruelle scattering theory do not overlap in coordinate space (cf. Duncan 2012, p. 276). Based on the results of the Haag-Ruelle theory, the LSZ formalism has been developed as a method more suitable to do calculations of actual processes. Part of the LSZ formalism is the LSZ reduction formula that is now "famous (and indispensable)" (Duncan 2012, p. 283) in QFT (cf. formula 5.3.34 above).

How far can these two work-arounds of Haag's theorem help to solve the two problems found in the last section? First of all, I wish to highlight that the Haag-Ruelle theory seems to answer both problems; it is mathematically rigorous, circumvents Haag's theorem, and its formalism is sufficiently close to the interaction picture to allow for the same causal interpretation. However, since my presentation of QFT in section 5.3 was given in the interaction picture with low and high energy cutoffs and perturbation theory, and that is what most physicists actually work with, it is important to have a closer look at this approach to QFT.[4] Introducing high and low energy cutoffs does not make invalid any of the points I made in sections 5.3.2 to 5.4. Thus, if QFT can be interpreted as describing causal interactions as here proposed, then taking this route around Haag's theorem will not change the situation for this interpretation. On the other hand, there is the problem of whether the interaction picture is suitable for realistic interpretations at all, given that it does not strictly follow from a set of coherent axioms. For example, Fraser (2006, p. 172) accuses QFT with cutoffs of "cheating", by using a mathematically ill defined apparatus. This feeling of discomfort is put into arguments by Fraser, claiming that QFT with cutoffs should not be taken as a basis for interpretations by philosophers, because it allegedly is inconsistent, underdetermined by evidence and makes unfeasible idealisations. But what is it exactly that makes a mathematically not rigorous theory unsuitable for interpretation?

Fraser (2009) makes the following argument for inconsistency. According to Haag's theorem a theory that accepts a certain set of principles T describes necessarily a free system. But the interaction picture is based on T and also incorporates the additional assumption that the system described is not free. Thus, there is an inconsistency in the interaction picture between T and the additional assumption: "Haag's theorem demonstrates that the interaction picture adopts an inconsistent set of principles" (Fraser 2009, p. 546 f.). Arguably, this is very bad news for the

[4] To clarify my terminology: I use the abbreviation 'QFT' for the formalism that I presented in section 5.3, including the renormalisation technique of introducing infrared and ultraviolet cutoffs. For better clarity I will sometimes explicitly mention that cutoffs are included. Some authors call this formalism 'Lagrangian QFT', 'conventional QFT' or 'QFT with cutoffs'.

philosopher of physics, because "[a] consistent set of theoretical principles may all be true simultaneously; an inconsistent set is certainly not." (Fraser 2009, p. 553)[5] However, as discussed above, Haag's theorem cannot be established if cutoffs are introduced. Thus, Fraser's argument for inconsistency cannot get off the ground in that case.[6]

In later papers, however, Fraser seems to put more weight on other arguments against QFT. In her (2011) paper, she argues that QFT and algebraic QFT (AQFT) are distinct though empirically equivalent theories, and that AQFT should be preferred. AQFT does not introduce cutoffs, and therefore is applicable in particular on arbitrary short length scales. On the other hand, given that today we can only test theories empirically on length scales far bigger than where the cutoff is applied, Fraser concedes that QFT and AQFT are empirically equivalent, and that on the basis of empirical evidence we cannot decide between them. Yet, Fraser argues that AQFT should be preferred for two reasons. First, only in AQFT important derivations like the spin statistics theorem and the CPT theorem are possible. Second, the cutoffs cannot be interpreted literally, since they would mean that we live in "a world in which space is discrete and of finite extent" (Fraser 2009, p. 551), a view that Fraser sees no support for. Thus, since AQFT does not include the cutoffs, Fraser seems to conclude that a discrete and finite space poses an unnecessary and unfeasible idealisation.

To my mind, Wallace (2011) has successfully rebutted Fraser's arguments. As Wallace highlights, QFT is "the great success story of post-war theoretical physics" (Wallace 2011, p. 116) with numerous new predictions that so far have stood up well to all empirical tests. Hence, QFT is a prime example of a physical theory to which the no-miracles argument for scientific realism can be applied. It seems very reasonable to expect that QFT is approximately true. On the contrary, AQFT so far has neither lead to any description of a realistic interacting system nor to any new prediction. Wallace (2011, p. 121) concludes from here that Fraser's underdetermination argument fails; there are not two empirically indistinguishable theories, since AQFT cannot be used to describe realistic interacting systems, there is only QFT. Furthermore, Wallace argues that also Fraser's other points can be doubted. First of all, the spin statistics theorem and the CPT theorem can be derived in QFT as well. Additionally, Wallace questions Fraser's claim that the cutoffs cannot be interpreted realistically.[7] There is no reason to expect that a theory of the arbitrary small does exist, and the success of QFT with cutoffs seems to be a

[5]A way to attack Fraser's argument from inconsistency, different form the one I present, could be to question her notion of 'inconsistency'. As Vickers (2013) argues, it should be avoided to talk of the inconsistency of *theories*, and not all inconsistencies are vicious. I cannot discuss here, whether this response would be successful.

[6]Fraser (2009) seems to dismiss QFT with cutoffs on the grounds that it is not Poincaré invariant. However, as Duncan (2012) shows, Poincaré invariance can be restored after the cutoffs have been introduced.

[7]Precisely, Wallace only refers to the cutoff of small distances, but I assume that his reasoning can equally be applied to large distances, assuming that there are good reasons to doubt that quantum fields extend over all of space-time.

reason to doubt that it does (Wallace 2011, p. 120). Even if it should eventually turn out that there is some physics below the scale where the cutoff is applied, we can expect QFT to be a good approximation of the new theory at length scales to which we now have experimental access. Finally, Wallace responds to the widespread resentment against QFT as being mathematically not well defined:

> Although the mathematical and physical pathologies of pre-1970s QED made this a reasonable objection prior to the development of renormalization-group methods, things have now changed, and the level of mathematical rigour in CQFT [i.e. conventional QFT] is basically the same as elsewhere in theoretical physics. (Wallace 2011, p. 124)

Therefore, rejecting QFT on the basis of a lack of mathematical rigorousness would amount to rejecting most of modern physics. According to Wallace, regarding AQFT as the mathematically rigorous rival of QFT is an outdated view.[8] I wish to add that the development of Haag-Ruelle scattering theory should also ease all those, who wish for more mathematical accuracy.

To be clear, the above is only meant to be a rough reconstruction of the debate mostly between Fraser and Wallace, and there are many subtleties that I cannot mention due to restrictions of space. I also do not wish to argue for the claim that interpreting AQFT should not be done at all. However, I hope that it has become clear why the philosophical account of the present work relies on a version of QFT that circumvents Haag's theorem by introducing high and low energy cutoffs.

In conclusion, I have argued that this theory can first of all stand up to the worry of lacking mathematical rigorousness, and second is suitable for the causal interpretation. It avoids the implication of Haag's theorem, and thus describes the continuous evolution of interacting fields. Finally, Haag-Ruelle scattering theory shows that it is possible to construct a mathematically rigorous theory of interactions.

6.3 The measurement problem

6.3.1 Introduction

The measurement problem is prima facie a general problem for any interpretation of quantum physics, which already might be a sufficient justification to discuss it here. Having said that, below I will motivate the particular relevance for the present work. Since the recognition of the measurement problem, continuous work by philosophers and physicists has brought no satisfying answer. All the different attempts to its solution face their own well-known problems. The general strategy in this chapter will be to first examine a formulation of the measurement problem that is supposed to show its insolubility by physics, and to argue that it is not convincing. Nevertheless,

[8] Another point of critique that is often put forward against QFT is its use of perturbative methods. This topic is not central for the discussion here, and I therefore simply wish to highlight that, ironically, perturbative methods have recently introduced into AQFT, in order to come closer to extracting any realistic models from it (cf. Duetsch & Fredenhagen 2000).

I will subsequently argue that the measurement problem remains significant in a weaker formulation. This weak measurement problem, however, can at present only be vaguely formulated and needs more work, done by physicists and not philosophers, so I will claim.

I wish to stress that I will not present a solution to the measurement problem, neither for the strong, nor the weak formulation. Rather, I will motivate a sidestepping strategy, that is, on the one hand I will acknowledge that the measurement problem is still open, on the other hand, I will argue that the present project can be pursued despite that fact. I have to admit that the result of my discussion will be inconclusive in that I cannot give any definite answers to the questions raised, but merely make suggestions. This might be too weak for some, however, I have to leave that for others to decide.

My main concern is towards two questions. 1) Is there any hope that the measurement problem can be solved by physics alone? 2) How can the philosophical interpretation of QFT proceed in the light of the measurement problem? Both questions are related, of course. Common proposals for the solution of the measurement problem are philosophical, rather than physical solutions, in the sense that none of them has any experimentally verified consequences and each one brings certain ontological commitments with it. Therefore, the general interpretation of QFT is dependent on which stance one takes towards the measurement problem, even though there are some topics in the philosophy of physics that are independent of it. In particular, if one is *not* willing to follow any of the common roads (Ghirardi-Rimini-Weber (GRW), Decoherence-Everett, De Broglie-Bohm) then, or so it seems, the interpretation of QFT cannot even start (cf. Barrett 2000; Wallace 2008, p. 16).

This has further implications for the theory of causation that I want to defend. For example, the De Broglie-Bohm theory in its current formulation is not covariant. Adopting it would place serious doubt on the STR, and therefore make various arguments that I present invalid. In the following, I will try to find a way out of this dilemma and argue that QFT can reasonably be interpreted, even though the measurement problem is not solved. To be clear, the task is not to solve the measurement problem, but to analyse whether a solution can be found in physics and whether it has to be solved in order to make the study of QFT worthwhile.

Central for my argument will be the distinction between subjective and objective solutions to the measurement problem. An objective solution would be a fundamental process in nature that somehow produces definite measurement outcomes. A subjective solution on the other hand only explains our experience of definite outcomes of measurements. As Schlosshauer (2004, p. 1271) puts it:

> In general, (macroscopic) definiteness—and thus a solution to the problem of outcomes in the theory of quantum measurement—can be achieved either on an ontological (objective) or an observational (subjective) level. Objective definiteness aims at ensuring "actual" definiteness in the macroscopic realm, whereas subjective definiteness only attempts to explain why the macroscopic world appears to be definite—and thus does not make any claims about definiteness of the underlying physical reality (whatever this reality might be).

It is crucial that 'subjective' does not necessarily mean something like 'caused by the subject' or 'only within the mind of the subject'. Rather, as it will turn out, the subjectivity is grounded in the fact that certain phenomena are relevant and only accessible from a specific location in the world. The names 'objective' and 'subjective' are to some extent misleading, as they do not necessarily exclude each other. I would therefore like to call these solutions *local* in opposition to *global* solutions, where the latter do not depend on a specific location. I will also slightly disagree with the view of Schlosshauer that subjective, or local, solutions are irrelevant on the ontological level. Even if there are no definite measurement outcomes, it might well turn out that our experience of definite measurement outcomes is grounded in real traits of the world and therefore has ontological significance. As a final remark to motivate this distinction, the discussion has shown that there are both potentially feasible global and local solutions on the table, with GRW and De Broglie-Bohm on the global and Decoherence-Everett on the local side.

The approach I take to weaken the measurement problem rests mainly on non-unitary evolutions. There might be other ways of potentially solving the measurement problem within standard quantum theory[9], for which non-unitarity is secondary or not important at all, thus making all what I say in the following sections obsolete. Landsman & Reuvers (2013) is a promising example for such an approach. I do freely admit that the latter might be the case, since this would only be grist to my mills by showing that the measurement problem can be solved without empirically unwarranted modifications of standard quantum theory or contentious interpretations thereof, and that more research in physics, not philosophy, has to be done.

6.3.2 The unsolvable problem

Even though what I am concerned with in this chapter is called the measurement problem, it is almost a consensus that the term 'measurement' should not appear as a primitive unanalysable term in QM. As Bell (1987, 1990) explains, first of all it is unclear what the term 'measurement' means. (What constitutes a measurement? Is it complete when an experiment gives a certain result or when the physicist is conscious of the result?) Second, if physics is a precise description of the microphysical world and an explanation of the observable phenomena, then it is unacceptable to have measurements as a black box within the theory. Leaving measurements unanalysed by physics not only neglects that the measured object and the measurement apparatus are made up of elementary particles, it amounts to acknowledging that there is something in nature of which we do not know what it is. This stands in extreme contrast to the methodological doctrine of physics to give a complete description of the physical world. Bub & Pitowsky (2010) critically call this analysability criterion for measurements a "dogma of quantum mechanics". However, I do not agree with the 'dogma charge', since, as explained, the criterion rests on at least two good motivations. But how far can the term 'measurement' be reduced and clarified in

[9]Given its wide acceptance, I count decoherence as a part of standard quantum theory.

QM?

To begin with, a measurement can roughly be defined as follows: A measurement is an interaction between object and apparatus by which information about the object before the interaction can be obtained from information about the apparatus after the interaction (cf. Wigner 1983, p. 329; Fine 1969, p. 111). To make this more precise, let us see how an interaction that can count as a measurement is described in QM. Suppose we have an object in state $|\psi_n\rangle$ that is an eigenstate of observable $\hat{A}$. Second, there is an apparatus that initially is in a state $|\phi_r\rangle$, where it is ready to measure the object's state. The apparatus does so by taking on one of its pointer states $|\phi_n\rangle$ that are correlated to the possible states of the object and eigenstates of observable $\hat{B}$. To let the systems interact with each other, they have to be coupled to $|\Psi\rangle = |\psi_n\rangle\,|\phi_r\rangle$. Now a unitary operator $\hat{O}$ can be found that makes the transformation $\hat{O}(|\psi_n\rangle\,|\phi_r\rangle) = |\psi_n\rangle\,|\phi_n\rangle$. This state is an eigenstate of the operator $\hat{I} \otimes \hat{B}$, with $\hat{I}$ being the identity-operator acting on $|\psi_n\rangle$, and the apparatus can be understood as measuring a property of the object (cf. von Neumann 1932, p. 235 f.).

So far there is no problem in this analysis. The measurement problem arises when the initial state of the object is not an eigenstate but a superposition of eigenstates $|\psi\rangle = \sum_n c_n\,|\psi_n\rangle$, with $n \geq 2$. The coupled system of object and apparatus will then evolve according to $\hat{O}[(\sum_n c_n\,|\psi_n\rangle)\,|\phi_r\rangle] = \sum_n c_n\,|\psi_n\rangle\,|\phi_n\rangle$, which is not an eigenstate of $\hat{I} \otimes \hat{B}$. The apparatus is now in a superposed state and the superposition has 'spread' from the object onto the apparatus. There is no unitary operator that would continue to evolve the superposition in such a way that the apparatus changes into a definite state $|\phi_n\rangle$. Furthermore, simply postulating that the superposition collapses into a definite state is not only *ad hoc* and in contradiction with the Schrödinger equation, but it is also highly unclear at what point in time this collapse should happen. Unfortunately, all this is contrary to our experience that macro physical objects always are in a definite state and not a superposition. With this background, the measurement problem can be formulated.

I want to distinguish between two versions of the measurement problem, a weak and a strong one. The weak version goes as follows (cf. Lyre 2010, sec. 2):

Weak measurement problem: How is a superposition of object states connected to the definite state of the apparatus that is observed?

This captures already the situation sketched in the previous paragraph, namely, that there has to be an explanation of why measurement devices always are in a definite state even though QM predicts that they are not. It seems, however, that the measurement problem can be formulated in a way that proves its insolubility by QM alone. This will be the strong measurement problem. It is one of the aims of this chapter to show that at the moment the weak version is preferable.

Two versions of the insolubility proof can be distinguished, which differ in respect of one of their premises. In one version the measurement problem is solved, if the final apparatus state is a mixed state constituted of eigenstates each weighted with a certain probability p.[10] For instance, the final apparatus state of a Stern-Gerlach experiment could be a mixture of pointer-state spin-up and spin-down, each with $p = 1/2$. This amounts solely to a disappearance of interference terms of the object-apparatus state. I do not hold this premise to be strong enough to solve the measurement problem for the following reason. One might argue in favour of an ignorance interpretation towards mixed states. That is, a mixed state merely shows our incomplete knowledge of reality, while the system actually is in only one of the components of the mixed state. However, this stance is already highly problematic concerning microphysical objects and much more so concerning a macro physical measurement apparatus. It is unclear what an apparatus in a mixed state would look like, since an experimental physicist in principal always knows the state of the apparatus. Additionally, "if the state of the apparatus which carried out the primary measurement is just as difficult to ascertain as the state of the object, it is not very realistic to say that the establishment of a correlation between its and the object's state is a fully completed measurement." (Wigner, 1983, p. 332) Therefore, the measurement problem is not solved if the apparatus remains in a mixed state and I opt for the other version of insolubility proofs. These have the premise that the measurement problem is only solved if the final state of the apparatus is a pure eigenstate of the measured observable.[11] In a non technical formulation the insolubility proof, which I call the strong measurement problem, goes as follows (cf. Albert 1994, pp. 73-79; Maudlin 1995, p. 7; Busch et al. 1996, pp. 91-93):

Strong measurement problem:

Premise 1: QM is complete and in particular the state vector is a complete description of a system.

Premise 2: The state vector always evolves according to linear and unitary dynamics.

Conclusion: A collapse of a superposition into a definite state cannot be described within QM.

It is worth pointing out that in some formulations (e.g. Maudlin 1995) premise two is stated in terms somewhat like this: 'The state vector always evolves according to the Schrödinger equation.' This, even though it is correct, is not helpful, since it conceals the real source of the measurement problem. This lesson can be learned from the insolubility proofs: Due to the *linearity* of the Schrödinger equation superpositions are possible. However, the main problem is that as a consequence of the *unitarity* of the Schrödinger equation a superposition will not evolve into an eigenstate. If the

[10] This argument can be found, e.g., in D'Espagnat (1966); Earman & Shimony (1968); Fine (1969, 1970) and Brown (1986). It is clear, though, that these authors are not claiming to solve the measurement problem, but merely use mixed final states as a premise to show the insolubility of the measurement problem.

[11] This argument can be found, e.g., in von Neumann (1932) and Wigner (1983).

premises are valid, the conclusion is that the measurement problem is unsolvable by QM alone.[12]

There is a second problem that has to be solved for a complete description of how measurement results come about. It has never been in the focus of research and I will only mention it briefly as well. It is the *problem of the preferred basis*: Every state $|\psi\rangle$ in QM can be represented by using different basis vectors, i.e., $|\psi\rangle = \sum c_i |\phi_i\rangle$. For example, a state that represents spin in x-direction can also be expressed using the spin in y-direction. Since both states have the same measurement probabilities, it is unclear which property actually is measured in an experiment (cf. Redhead 1987, p. 57; Schlosshauer 2004, p. 1270; Breuer & Petruccione 2002, p. 271 f.). It also follows that any apparatus that can measure one observable can in fact measure every observable. This is an absurd consequence given the technical limits of real measurement devices (cf. Breuer & Petruccione 2002, p. 273). However, the strong measurement problem can be considered to be more severe and I will concentrate on it. In the remainder of this chapter, I will not follow one of the common roads (i.e. GRW, Decoherence-Everett or De Broglie-Bohm), instead, I will attack premise two of the strong measurement problem.

6.3.3 Decoherence as a non-unitary evolution

In the last four decades it has become clear that the theoretical description of a measurement, outlined above, cannot be considered complete. It neglects that the object-apparatus system is never totally isolated, but always interacts with its environment. This failure is removed by decoherence, which in recent years has been experimentally verified.[13] The purpose of this section is to show that decoherence is a non unitary process that has to be taken seriously, that is, it cannot completely be dismissed as having no implications for the measurement process. If this is true, then premise two of the strong measurement problem is false, since the state vector does not always evolve according to unitary dynamics. As a consequence, there is no proof of the insolubility of the measurement problem. Even though I want to stress that this does not constitute a solution to the measurement problem either.

To gain a better idea of how decoherence works, let us take a look at a simple model that can be found in Zurek (1982, 2003), Giulini et al. (1996, ch. 3.1.3) and Schlosshauer (2004, pp. 1276-1278). Suppose that we have a system $|\psi\rangle$ that is surrounded by an environment $|E\rangle$. For simplicity, the object, which is measured, and the apparatus, which measures the object, are combined in the system $|\psi\rangle$. Writing them separately would not essentially change the model. Now, at time $t = 0$ the system is in a superposition of spin states

[12] Maudlin formulates this measurement problem in three statements that taken together are inconsistent and therefore at least one of them has to be false. This is certainly a valid formulation; however, when we ask why the statements are inconsistent we have to come back to the insolubility proof that therefore lies at the heart of the problem.

[13] See the results described in Brune et al. (1996), Breuer & Petruccione (2002, ch. 4.5) and Hornberger et al. (2003).

$$|\psi(t=0)\rangle = a\,|\uparrow\rangle + b\,|\downarrow\rangle\,.$$

Since the system is not isolated, it will soon become entangled with its environment and the state will read as

$$|\psi(t>0)\rangle = a\,|\uparrow\rangle\,|E_{\uparrow}\rangle + b\,|\downarrow\rangle\,|E_{\downarrow}\rangle\,.$$

The effect of decoherence can be best seen when looking at the reduced density matrix of the system

$$\begin{aligned}\rho_S(t>0) &= Tr_E[|\psi(t>0)\rangle\,\langle\psi(t>0)|]\\ &= |a^2|\,|\uparrow\rangle\,\langle\uparrow| + |b^2|\,|\downarrow\rangle\,\langle\downarrow| + z(t)ab^*\,|\uparrow\rangle\,\langle\downarrow| + z^*(t)a^*b\,|\downarrow\rangle\,\langle\uparrow|\,.\end{aligned}$$

Here the time dependence of the off-diagonal terms is given by the function $z(t)$. It is the overlap of the states of the environment $|E_n(t)\rangle$ and $|E_m(t)\rangle$ and generally is dominated by an exponential function

$$|\,\langle E_n(t)|\,|E_m(t)\rangle\,| = exp[\Gamma_{nm}(t)],\ \Gamma_{nm}(t) \leq 0.$$

As can be shown, "[t]he time dependence of the decoherence function $\Gamma_{nm}(t)$ strongly depends, in general, on the specific form of the system-[environment] coupling, on the various parameters of the underlying microscopic model, and also on the properties of the initial state." (Breuer & Petruccione 2002, p. 222) Furthermore, in realistic models the decoherence function will lead to a fast decay of the off-diagonal terms of the system's density matrix. So after some more time the latter is given as

$$\rho_S(t \gg 0) \simeq |a^2|\,|\uparrow\rangle\,\langle\uparrow| + |b^2|\,|\downarrow\rangle\,\langle\downarrow|\,.$$

The density matrix has no interference terms and is therefore equal to the density matrix of a mixed state. This is one of the main consequences of decoherence, i.e., the cancelling of interference terms in the system's reduced density matrix. Since, however, decoherence does not lead to a density matrix of a pure and non-superposed state, it is clear that the measurement problem cannot be solved by decoherence alone.[14]

Even though decoherence does not solve the measurement problem, it is significant that it involves a non-unitary evolution of the systems state. This can be seen for instance by looking at the expectation values for a suitable observable for the spin state of the system. If this observable is $\hat{S} = \hat{S}_S \otimes \hat{I}_E$, where $\hat{I}_E$ is the identity operator for the environment, it is clear that

[14]There are two interesting points to note: (1) Decoherence is in principle reversible, because the involved interaction Hamiltonians are, at least in some models, hermitian (cf. Zurek 1982, p. 1873, Giulini et al. 1996, p. 23, 45) and (2) the interference terms do not always vanish completely in finite time and could be measured in a suitable experiment, if we had the necessary technology (cf. Giulini et al. 1996, p. 54).

$$\langle\psi(t>0)|\hat{S}|\psi(t>0)\rangle \neq \langle\psi(t\gg 0)|\hat{S}|\psi(t\gg 0)\rangle$$

or equivalently

$$Tr_S\left\{\hat{S}\rho_S(t>0)\right\} \neq Tr_S\left\{\hat{S}\rho_S(t\gg 0)\right\}$$

and therefore the criterion for non-unitarity is fulfilled.

In more general terms, that is, not in direct relation to the above simple model for decoherence, the non-unitarity of the system's evolution can be seen from the following. If the density matrix of a system behaves unitarily, its time evolution is determined by the von Neumann equation

$$i\frac{d\rho}{dt} = \left[\hat{H}, \rho\right].$$

This equation is equivalent to the Schrödinger equation (either one can be derived from the other), which determines the time evolution of the state's wave function. In contrast to this, if decoherence is present, most models describe the time evolution of the system's reduced density matrix by a master equation of the Lindblad form (cf. Giulini et al. 1996, pp. 111-112; Breuer & Petruccione 2002, pp. 119-222):

$$i\frac{d\rho_S}{dt} = \left[\hat{H}, \rho_S\right] - \sum_{i=1}^{N^2-1} c_i \left(\hat{A}_i\rho_S\hat{A}_i^\dagger - \frac{1}{2}\hat{A}_i^\dagger\hat{A}_i\rho_S - \frac{1}{2}\rho_S\hat{A}_i^\dagger\hat{A}_i\right).$$

The second term on the right hand side is the non-unitary part of the equation and implies a non-unitary evolution of the system's wave function.[15]

It might be objected to the above point that the attribution of a state to the system alone is not possible, because it is entangled with its environment and the decisive feature of entanglement is non-separability. To answer this objection, we have to enter the ongoing discussion about entanglement measures and criteria for separability. This is already a large and fast growing field, making it impossible to give a complete survey here.[16] As the situation is today, there is no single measure for entanglement, applicable and useful in all situations, and no single criterion for separability, giving sufficient and necessary conditions for all possible states. I will therefore only discuss the most common measure, which gives the result a somewhat limited scope.

The oldest criterion for separability and at the same time measure for entanglement is given by the Bell inequalities. Though it has been shown that not all entangled states violate them, those that do not can possibly be transferred into states that do. Bell inequalities therefore represent a powerful tool to detect entanglement (cf. Munro et al. 2001). They are also intuitively very accessible, since they can

[15] The description of decoherence using a master equation has interestingly the same form as statistical collapse models, like that of GRW, have (cf. Giulini et al. 1996, p. 68 f., 223 and ch. 8).

[16] For review articles see Horodecki et al. (2009); Plenio & Virmani (2007) and Li et al. (2010).

be tested in experiments.[17] The connection to decoherence is now straightforward. The strength of violation of the Bell inequalities is determined by the interference between the entangled states.[18] In particular, the Bell inequalities are not violated, if there is no interference. Since decoherence eliminates interference, it is clear that decohered states will satisfy the Bell inequalities and therefore are not entangled (cf. Bertlmann et al. 2002, ch. 11.6; Bertlmann et al. 2003). More precisely, since decoherence becomes stronger over time, a "time of disentanglement" (Blanchard et al. 2001) can be defined, after which a previously entangled state is separable. This influence of decoherence on entanglement has also been experimentally confirmed by de Riedmatten et al. (2006).[19]

However, it has to be stressed that decoherence does not erase all correlation between system and environment. This is not possible, since the evolution of ρ_S clearly depends on that of the environment it is coupled to. Rather, decoherence "retains all 'statistical' correlations [...] while dropping all quantum correlations (entanglement)." (Zeh 2000, p. 22) The system is therefore often called an 'open quantum system'. Again, it has to be stressed that the Bell inequalities are not sufficient and necessary to show the separability of fully decohered states in all cases. Additionally, the connections between different measures of entanglement and between different separability criteria are not entirely clear. Nevertheless, the satisfaction of the Bell inequalities alone already demonstrates that fully decohered states have no non-local features in experiments and thus cannot be genuinely entangled.

I now want to discuss the relevance of decoherence to the measurement problem. The theoretical description of decoherence, adumbrated above, rests on the trace operation, which is on the surface just an ad hoc manipulation of the theory; we are only interested in the degrees of freedom of the system and therefore we simply delete the degrees of freedom of the environment. However, this simplification of the actual physical situation is justified by the perspective of the local observer, who, if the Hilbert space of the total system is composed of the Hilbert spaces of the system and the environment, $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_E$, can be fully described within $\mathcal{H}_S$. If we could change from the local into a global perspective, then we would see that there is no decoherence and that the entanglement between system and environment remains intact. This is the reason why most, if not all, authors do not count decoherence as one of the fundamental processes in nature.

Should we therefore discard decoherence completely and not use it in any explanation of how measurement results come about? I think not. We have to be clear about what is doing the work in the explanation of phenomena and experimental

---

[17] Today, mostly the more refined method of "entanglement witnesses" is used to detect entanglement, which can be considered, loosely speaking, to be a generalisation of the Bell inequalities (cf. Horodecki et al. 2009, ch. VI.5). Bell inequalities also not only detect, but also measure entanglement by their connection to the concurrence measure (cf. Emary & Beenakker 2004).

[18] Kiess et al. (1995) have performed an experiment in which they manipulated the interference term to gain different degrees of violation of the Bell inequalities.

[19] See also the survey of experiments in Schlosshauer (2007, ch. 6).

results involving decoherence. The explanatory work is not only done by the local perspective of the observer, but also by the underlying constitution of the world that makes the description using decoherence possible. The world could be different in a way that does not incorporate decoherence, hence the fact that we can use decoherence must be grounded in some feature of the world. This situation is analogous to other phenomena that involve dissipation. For example, the loss of speed of a tyre rolling over the floor can be explained by friction, even though it is clear that friction is just a simplified way to describe more fundamental forces. One might object, however, that the above analogy suggests that the non-unitarity of decoherence might be reduced to processes in nature that are in fact unitary.

At this point the distinction between local and global (or subjective and objective) solutions to the measurement problem becomes important. The measurement problem points towards the incapability of physics to explain definite measurement outcomes. However, in so far measurement outcomes are local events and always observed from a local perspective, a local solution to the measurement problem would be sufficient. For this reason we can make use of local phenomena like the non-unitarity of decoherence in explicating measurement outcomes.

Schlosshauer (2004, p. 1271) calls such a solution 'subjective', because it "does not make any claims about definiteness of the underlying physical reality". I disagree here, because even though decoherence disappears globally, the fact that we locally observe decoherence clearly tells us something about the underlying physics. Unitary physics can have effects such that for the local observer they are non-unitary. The correlation of a system with a non-local environment has the effect that a certain property of the system, its coherence, is not observable locally anymore (cf. Hornberger 2009, p. 226). The actual justification for taking decoherence as a real process is of course the experimental results. Thus, decoherence is the description of a real process, the effects of which can only be observed locally. For these reasons I believe that decoherence and its non-unitarity can be relevant to explain definite measurement outcomes, despite being only a local phenomenon.

Even though the non-unitarity of decoherence is mentioned by several authors[20] it has received little attention in the discussion about the measurement problem.[21] The reason is probably that decoherence alone does not solve the measurement problem, and to be clear, I will not change this situation. Nevertheless, it is important to investigate the relevance of decoherence a bit further. Let us recall premise two of the strong measurement problem given above: 'The state vector always evolves according to linear and unitary dynamics.' Given that my argument in favour of local solution is correct, this premise can be given a local form: 'The state vector of a *local* system always evolves according to linear and unitary dynamics.' Since this is a statement about the dynamics of a system and decoherence adds something new

[20]See for example Busch et al. (1996, p. 123), Giulini et al. (1996, p. 51, 249), Kupsch (2000, p. 128) and Breuer & Petruccione (2002, p. 109, 122).

[21]Much more attention to the non-unitarity has been paid in the field of quantum computation, where it is particularly problematic, since a non-unitary evolution typically goes hand in hand with a loss of information (cf. Facchi et al. 2005).

to the dynamics, clearly, the non-unitary evolution of the decohered state has to be added to premise two. It is obvious that the non-unitarity cannot be added by a conjunction, because this would make premise two contradictory; the state vector cannot *always* evolve according to unitary dynamics and sometimes not. So we have to try a disjunction, and since unitary and non-unitary dynamics cannot happen at the same time, it has to be an exclusive disjunction. Premise two then reads: 'Either the state vector of a local system evolves according to linear and unitary dynamics or according to non-unitary dynamics.'

The premise is fine now, but what about the conclusion? Does it still follow? Here we have to keep in mind that the strong measurement problem is just a vague translation of what really is a precise mathematical proof. What the proof sets out for is to show the insolubility of the measurement problem within standard QM. Since unitarity of the evolution is ubiquitous to establish the proof, it is clear that the conclusion does not follow from premise one with the new premise two. It might be argued that still at least sometimes the state vector evolves according to a standard Schrödinger equation and therefore at least in these cases the insolubility proof is correct. However, in all interesting cases, those that can count as measurement, decoherence and therefore non-unitary dynamics are effective and the conclusion does not follow.

The result of all this is that, since decoherence is part of every measurement from a local perspective, the insolubility of the measurement problem cannot be proven. The strong measurement problem is only retained if a global solution is demanded, but is false if a local solution suffices. However, it is most important to stress that decoherence does not solve the measurement problem, since it does not explain how a superposition can evolve into a definite state, given that the outcome of the decoherence process is a mixed state. The weak measurement problem is still unsolved. Nevertheless, the situation has changed considerably. Was it hopeless before to look for a solution to the measurement problem in ordinary physics, this is now a possibility. It cannot be strictly excluded that a state evolves according to non-unitary dynamics that could imply the collapse of a superposition into a definite state. Of course, from the fact that decoherence is non-unitary, it does not follow that there must be other (non-unitary) processes that actually solve the measurement problem. The point I wish to make by involving decoherence is merely to cut the chains in which the strong measurement problem has put the philosopher of physics, and thereby at least take the possibility of a physical solution into account. In the next section, I will look into QFT to see whether there are hints that could eventually lead to a solution of the measurement problem.

6.3.4 Can quantum field theory help?

By the fact that QFT, as it is found in textbooks, only involves unitary dynamics, it seems that QFT does nothing to clarify the measurement problem in addition to QM. This opinion is for example endorsed by Barrett (2000, p. 4 f.) and Wallace (2008, p. 83). The present section is a critical assessment of this view. To begin with, it will be instructive to take a look on how measurements function in particle physics from the viewpoint of QFT.

The first argument for the relevance of QFT in measurements comes from experimental practice. Fundamental particles can only be measured in specially constructed detectors. I have already described the experimental basics of measurements in section 5.2.3 and only want to summarise the main points again. The condition sine qua non for measurement is that energy goes from the measured particle onto the detector. This is usually reached via two processes: ionisation or emission of radiation. Properties that are measured are usually time, momentum, velocity or total energy. Furthermore, every measurement of these properties can automatically be used to measure position; simply given by the fact that the detector covers a certain area in space and that the particle is in that area when it is detected. The accuracy of the position measurement depends on how small the parts can be made that a detector is composed from. In practice, these processes are described through approximations derived from an exact treatment in QFT. In the end, however, every interaction of a particle with a detector is a scattering event that can be described only in QFT, and not QM.

In comparing QM with QFT there is yet another interesting fact that is sometimes noted as a curiosity. While in QM the wave function evolves deterministically, measurement results can only be predicted with a certain probability, following from the Born rule. In QFT on the other hand, interactions or decays are the place where probabilities occur. Eugene Wigner expresses his surprise about probabilities in QM when he writes: "The place where one expects probability laws to prevail is the change of the system with time." (Wigner, 1983, p. 326) The application of the Born rule, however, cannot be regarded as describing the 'change of the system with time'. According to Cartwright (1978, p. 54), the received interpretation of $|\Psi(x)|^2$ is that it is the probability density that a particle is in a small region around x. Cartwright points out that this cannot be true literally, since experiments like the double slit experiment show that quantum mechanical particles are not located in that way. On the other hand, the usual alternative, to interpret $|\Psi(x)|^2$ as the probability for a particle to be at x when measured, is not feasible either since it leaves unclear what a measurement is and runs into the measurement problem (cf. Cartwright 1978, p. 55). Hence, the meaning of the probabilities remains unclear in QM. In the light of the foregoing discussion, however, it is hard to overlook the parallelism between QM and QFT. Measurements are nothing but interactions between different systems for which QFT assigns probabilities for an initial state to evolve into a final state. It is therefore tempting to identify the Born rule probabilities of QM with transition probabilities of QFT. This is exactly the programme that has been pursued by Cartwright (1978, 1980).[22]

Cartwright (1978, p. 55) proposes that

> $|\Psi(x)|^2$ does not represent a probability at all. Hence there is no problem of finding an event for it to be the probability of. [...] The only probabilities there

[22] See also Maxwell (1982, 1988) for a similar proposal.

> are in quantum mechanics are probabilities for energy interchanges (transition probabilities to energy eigenstates).

This is precisely a consequence of the fact that "[a] real detector cannot respond to the mere presence of a particle." (Cartwright 1980, p. 110) Real measurements only happen when a particle transfers energy to a detector, and this process is a quantum field theoretical scattering process that cannot be described in QM.

However, given that the Born rule is actually applicable and gives probabilities for a range of real processes, the above proposal faces what Cartwright calls the 'consistency problem'; it needs to be shown that the Born probabilities agree with the probabilities derived from QFT. More precisely, if we are not interested in the actual number of particles, then all what needs to be shown to answer the consistency problem is that relative probabilities in QM and scattering theory are equal, that is, the probability that a particle is (scattered) at x divided by the probability that it is (scattered) at x'. Indeed, Cartwright (1980) claims to have proven that this is the case. I cannot sufficiently discuss Cartwright's proof, and only wish to draw two conclusions. First, further work has to be done to solve the consistency problem and to show why the Born rule can give probabilities for real processes, a full description of which goes beyond the recourses of QM. Second, and more positively, it seems safe to say that in the discussions around the measurement problem the fact that "[t]he exchange of energy is the basic event that happens in quantum mechanics; and the basic event whose effects are theoretically described and predicted" (Cartwright 1978, p. 55) has been strongly under appreciated. I would therefore hope that more research in that direction will be done in the future.

Once it is agreed that measurement processes have to be described in QFT, the latter also seems to give hints for a possible solution of the measurement problem in quite another way than Cartwright's programme. If decoherence is effective in ordinary QM, it is much more so in QFT. There are two ways for decoherence to work in QFT, either the matter is decohered by a field or a field is decohered by the matter (Here 'matter' and 'field' are merely common physicist's jargon. One could also say 'matter field' and 'force field'). Which description is right depends on the actual situation that is described.[23] It is possible to find the reduced density matrix either for the matter or the fields and describe their dynamics. Therefore, in realistic situations it might be the case that superpositions, which cause the most trouble according to the received measurement problem, do not take part in scattering events, but that only mixed states do. On top of that, Hartle (1994) has shown that on the basis of decoherence it is possible to define non-unitary dynamics for interactions that respect the most basic features of QFT like energy conservation and no signalling faster than light. Though it remains to be seen, if these suggestions actually are a step in direction to a solution of the measurement problem.

Additionally, it is interesting that proposals for non-unitary dynamics are relatively widespread in relativistic quantum physics. Starting from Hawking (1982) they

[23]See Kiefer (1992) and the chapter by Kiefer in Giulini et al. (1996). See also chapter 12 in Breuer & Petruccione (2002) for a discussion of bremsstrahlung and decoherence.

are discussed in relation to black hole physics and quantum gravitation (cf. Banks et al. 1984; Srednicki 1993; Unruh & Wald 1995). As Hartle (1994, p. 34) remarks, "[w]hen spacetime geometry is not fixed, as in quantum gravity, or when it is fixed but not foliable by spacelike surfaces, some modification of familiar quantum theory seems inevitable." Furthermore, non-unitary dynamics are supposed to be the right description for the phenomenon of neutrino oscillation (cf. Goswami & Ota 2008), and important in the research on quantum dissipation, which is closely related to decoherence (cf. Garbaczewski & Olkiewicz 2002). I conclude from these ongoing research projects that philosophers should be more open towards non-unitary dynamics and investigate further the implications on the measurement problem.

In conclusion, QFT at the present stage does not solve the measurement problem, but at the same time there is no definite answer to the question whether QFT could solve it; more research in theoretical physics has to be done. Nonetheless, QFT certainly has implications for our understanding of the measurement process by making the latter much more lucid. Simply extending the insolubility proof of the measurement problem onto QFT, as it is found in textbooks, leads to an obstruction of one's ability to appreciate ongoing developments in physics that could be relevant for the measurement problem. However, given that all what has been said does not solve the measurement problem, in the next section I will investigate what this means for the present project.

6.3.5 Interpreting physics despite the measurement problem

What conclusions can be drawn from the previous two sections for the interpretation of QFT. Can this physical theory be interpreted despite the still unsolved measurement problem? The first result of the foregoing discussion is that the hope to solve the measurement problem within quantum physics, and not by a philosophical theory, is not completely illusory. The argument for the insolubility of the measurement problem in quantum physics rests on the premise that only unitary transformations occur. This argument, however, is only correct if a global solution is demanded and it is false for local solutions, because of decoherence. The latter is clearly relevant for all measurements, it is experimentally verified and it leads to a non-unitary evolution of quantum states. It is therefore not true that quantum states always evolve according to unitary dynamics and the possibility is open to solve the measurement problem by, e.g., finding special conditions within quantum physics, in analogy to decoherence, under which a superposed quantum state reduces to a definite state. Therefore, the conclusion of the strong measurement problem, as I named it above, does not follow. However, the weak measurement problem, that is, to find some description of how measurement results come about, remains. This leads is to QFT.

The credo of Maxwell and Bohm to ban the term 'measurement' from physics, or at least reduce it, cannot be given justice in QM, but quite possibly in QFT. It is clear that a measurement process is not fully captured within QM by taking the object state $|\psi_n\rangle$, the apparatus state $|\phi_r\rangle$, coupling both to the joint state $|\Psi\rangle = |\psi_n\rangle\,|\phi_r\rangle$ and transforming the apparatus to a pointer state $|\phi_r\rangle \to |\phi_n\rangle$. Measured objects undergo scattering processes in detectors that can only be described by QFT. Roughly, a measurement nowadays is just an ionisation or excitation interaction by which an electronic signal above a certain threshold can be created. Thus, QFT is a contender for what Maxwell (1976) calls a "micro realistic theory of measurements".

From the viewpoint of QFT, the description of a measurement in QM is not even a simplification, but at least in some cases clearly false. Measurement processes that involve the creation or annihilation of particles, as for instance all measurements of photons or energy do, cannot be described just by coupling the object to the apparatus. From this perspective it is astonishing how it can be expected to reach a full description of measurements, including a solution to the measurement problem, by using only the limited tools of QM. Applying QFT makes measurements a more complicated process, but at the same time opens up possibilities for new discoveries. To gain better understanding of the implications of QFT on the measurement problem, it certainly has to be investigated what happens with superpositions under scattering processes in detector materials. The results of such investigations remain to be seen, but the door is open for solving the measurement problem within QFT.

At this point two possible objections to my general argument in favour of QFT and the philosophical interpretation thereof have to be discussed. First, one might argue that I ignored a phenomenon that is equally important and problematic as the measurement problem and connected to the latter, namely entanglement and EPR-correlations that show up in experiments. Since I have been expressing the hope that a satisfying description of the measurement of superpositions can be found in QFT, it is reasonable to ask of me to explain how spacelike separated measurements of an entangled state can be correlated. A description of measurements in QFT that leaves out Bell-experiments cannot be considered to be complete. I will discuss entanglement in more detail in the next chapter and for now only have one brief comment. It is possible that a full quantum field theoretical description of measurements of superpositions has implication for Bell-experiments. The reason is the connection between the so-called no-signalling condition and linearity. If $|\Psi_{AB}\rangle$ is an entangled state of A and B with density matrix ρ_{AB} then measurements at all times made on system A can only depend on the reduced density matrix of A, if there is no interaction, i.e., signalling between A and B. The reduced density matrix is gained by the trace rule, $\rho_A = Tr_B \rho_{AB} = \{p_i, \hat{P}_{A_i}\}$, where the $\hat{P}_{A_i}$ are the projection operators for the states $|A_i\rangle$ and the p_i are the corresponding probabilities. Now, if $\hat{H}$ is the evolution operator, the no-signalling condition is equal to the linearity of $\hat{H}$, i.e.,

$$\hat{H}\left(\sum_i p_i \hat{P}_{A_i}\right) = \sum_i p_i \hat{H} \hat{P}_{A_i}.$$

In other words, the no-signalling condition is equal to the condition that the measurement statistics at A do not change over time, no matter what is done at B, and hence equal to the linearity of the time evolution for the density matrices. On the other hand, if non-linear dynamics should turn out to be important for measurements, as the dynamics for open quantum systems imply (cf. Breuer & Petruccione 2002, ch. 3), this could lead to a relaxation of the no-signalling condition (cf. Simon et al. 2001; Holman 2008). A full description of measurements therefore could have important consequences for the problem of EPR-correlations. For these reasons, EPR-correlations do not threaten the ontological interpretation of QFT additionally to the measurement problem.

The second objection is that even though the foregoing argumentation might be correct and it is not necessary to follow Decoherence-Everett, GRW or De Broglie-Bohm, it is certainly still possible to follow them, under the assumption that they work. So it might be argued that even now it is somehow more rational to take one of these three solutions instead of leaving the measurement problem unsolved. Though it is trivially true that it is better to have the measurement problem solved, I do not share this rationale. For one thing, it is clear that leaving the measurement problem open cannot be the final word, but adopting GRW, Decoherence-Everett or De Broglie-Bohm might not be any better. Discussing these positions is beyond the scope of this chapter, but it is well-known that all of them come only with a price. Which one is cheaper, GRW, Decoherence-Everett, De Broglie-Bohm or an open measurement problem is hard to estimate and quite possibly only a matter of taste. Furthermore, considering the overwhelming success that QFT has had, if there is just the slightest chance that a satisfying description of measurements can be found in QFT, it is not absurd to be optimistic and hope that QFT can overcome this problem rather than to break down.

However, the most compelling argument in my eyes comes from naturalism. This position roughly inherits that physics and not a priori philosophy is the way to show what the world is like. It goes without saying that therefore philosophers should follow physics as far as possible. If now it were true that the measurement problem is insoluble within physics, this would be the end of the physical road and the naturalist had to (at present) choose between GRW, Decoherence-Everett or De Broglie-Bohm. A decision between them cannot yet be grounded in physics, because neither one has any empirically verified consequences. However, since the insolubility proof is invalid, it seems to be more in the spirit of naturalism, so to speak, to hope for a physical solution of the measurement problem, rather than to abandon physics.

To conclude, I hope to have shown that it can be a fruitful task to interpret QFT, even though the measurement problem is still unsolved and that the measurement problem does not necessarily need a philosophical solution. Having said that, I want to stress the following. In this chapter I have pointed towards some very speculative ideas in physics, e.g. non-unitary evolutions, that, so my conjecture, might have some impact on the measurement problem. I do not endorse any of these speculations that go beyond the standard model. The point I wanted to make was merely to show that physics is not defeated yet, and that it is reasonable and in the spirit of naturalism, to hope for a physical solution to the measurement problem, rather than a philosophical one. In other words, by holding on to the strong measurement problem we rob ourselves of interesting possible solutions. The investigation into QFT also has shown that the usual description of measurements in QM is not only simplified, but also clearly false for most cases. However, for a clear account of how the measurement process functions and where exactly the problem is, the behaviour of superpositions in QFT has to be examined. More research in theoretical physics has to be done. Only if we have a precise and accurate question, an answer can be found.

6.4 Is entanglement a causal relation?

6.4.1 Entanglement

Within this section, I am concerned with the question whether correlations due to entanglement can be interpreted as causal relations. It is not my aim to provide any further understanding of entanglement, and many equally interesting questions will remain unmentioned. Some of my arguments concerning causation will rest on the validity of STR, which many regard as not given for entangled states. Consequently, within this subsection, I will first provide a general description of entanglement and then argue that it is not in conflict with STR.

Entanglement is a curious thing, especially when exploited in experiments of the type first proposed by Einstein et al. (1935). A paradigm example for entanglement is the singlet state of two electrons:

$$|\psi\rangle_{Sin} = \frac{1}{\sqrt{2}} \left(|{\downarrow}\rangle_a |{\uparrow}\rangle_b - |{\uparrow}\rangle_a |{\downarrow}\rangle_b \right) \tag{6.4.1}$$

Where $|{\downarrow}\rangle_a$ is an eigenstate and means that electron a has spin state down, and $|{\uparrow}\rangle_a$ means that electron a has spin state up (respectively for b) and only spins on one axis are regarded. In a classical EPR-type experiment a singlet state is produced and two measurements are performed on it, say, by Alice and Bob. Application of the Born rule now tells us that the probabilities for the first measurement, by Alice, are equally $1/2$ for measuring either spin up or spin down. The interesting trait of EPR-type experiments now is that whenever Alice measures spin up, the conditional probability for Bob to measure spin down is 1, and correspondingly when Alice measures spin down Bob will measure spin up. As experiments have shown, Alice and Bob can be timelike separated in their reference frames and there is no known physical signal connecting their experiments, but still the perfect correlation holds.

This, of course, was only a very rough description of what is seen in experiments. A more careful analysis of the situation can be given with the help of Bell inequalities. As Bell has proven, any theory that fulfils three conditions cannot possibly reproduce the correlations predicted by QM and found in experiments.[24] These three conditions are (cf. Berkovitz 2014):

Locality: Probabilities for each measurement are independent of each other. Or, in other words, "[t]he probability of joint outcomes is equal to the product of the

[24] It is sometimes claimed that less conditions are necessary, however, as Żukowski & Brukner (2014) show, this is not the case.

probabilities of the single outcomes." (Berkovitz 2014) This is often translated into the condition that whatever a measurement depends on, must lie in the backwards lightcone of that measurement. Furthermore, it can be shown that locality is equivalent to the conjunction of two conditions (cf. Butterfield 1992b, p. 27): 1. parameter independence, which means that Alice's measurement result is independent of Bob's detector setting (and vice versa), and 2. outcome independence, which means that Alice's measurement result is independent of Bob's measurement result (and vice versa).[25]

Independence: The complete distribution of possible states λ is independent of the detector settings. Or in other words, the state that is measured is independent of the devices that measure.

Realism: The quantum mechanical state $|\psi\rangle$ is an adequate description of the actual state λ of the system, even though the actual state might not be equivalent to $|\psi\rangle$. Thus, there might be hidden variables, which are not part of the formalism of textbook QM.[26] In particular, when $|\psi\rangle$ is an entangled state, hidden variables are necessary to formulate the locality condition (cf. Źukowski & Brukner 2014, p. 3).

Since experiments have repeatedly reproduced the predictions of QM, it follows that at least one of these three conditions has to be wrong.[27] Quite often, one finds the assertion that it must be locality that is violated by entanglement, which also seems to have been the opinion of Bell himself. While this might indeed be the most reasonable choice given some widely shared background assumptions, it is by no means certain, and prima facie any of the three conditions can be wrong (cf. Źukowski & Brukner 2014). Having said that, I will follow the standard approach here and assume that indeed the locality condition is violated. The next relevant question, as will become clear in the next section in particular for questions on causation, is whether it follows from a violation of locality that there is a conflict between QM and STR. I will now argue that this is not the case and that a no-signalling condition holds for entangled states.

The no-signalling theorem (NST) is a general constraint in relativistic QM that

[25] Some authors have a far more inclusive notion of 'locality' than the one endorsed here. Thus, when explicating the condition, one has to be careful to not conflate the terminology. Furthermore, the division into parameter and outcome independence has been called into question, e.g., by Maudlin (2002).

[26] See Źukowski & Brukner (2014, p. 3) for a list of what these hidden variables might be.

[27] To be more precise, all experimental tests of Bell inequalities that have been made require additional premises to the ones necessary to derive the Bell inequalities. Thus, a violation of Bell inequalities in these experiments could have been due to violations of these additional premises. Usually this is called the possibility of loopholes. Even though, each known loophole has been closed in separate experiments (cf. Larsson et al. 2014), as of today, no single experiment could be carried out that closes all loopholes simultaneously. Consequently, there is still the possibility that the violations of Bell inequalities comes from a loophole and not from a violation of the three condition necessary to derive Bell inequalities.

will be important for the argument of the next section.[28] In brief, the NST means the following:

No-signalling: "The statistics (relative frequencies) of measurement results of a quantummechanical observable cannot be altered by performing measurements at a distance." (Redhead 1987, p. 113)

The NST is motivated by, but does not follow from the relativistic constraint of Lorentz invariance, which in QM translates to the constraint that spacelike separated operators commute. However, it is a sufficient condition for the Lorentz invariance of relativistic quantum theories (cf. Healey 2014, p. 10), and it is therefore reasonable to assume that NST and Lorentz invariance stand and fall together. An immediate consequence of the NST is that entanglement cannot be exploited for sending messages. Whatever Bob does on his apparatus, it will not change the statistics of the measurement outcomes that Alice finds. Therefore, Alice cannot know how Bob has set his apparatus or whether Bob has performed a measurement at all. Even though the NST looks very similar to the locality condition, it is crucial to keep them separate; each can be true without the other (cf. Skyrms 1984, sec. V).

Having said that, the last claim is highly controversial and needs qualification. First of all, locality is about probabilities for single measurements, while the NST is about relative frequencies. Thus, it is possible that a failure of locality implies that a single measurement of Alice alters the probability of a single measurement result for Bob, but that this effect is not visible in the overall statistics (cf. Butterfield 1992a, p. 63). There is, however, the more difficult problem of what non-locality would imply for STR and the NST.

As Seevinck & Uffink (2011) argue, the locality condition of the Bell inequalities should be understood as motivated by and encoding the relativistic constraint that an event A can only directly be influenced by other events B that lie in A's backwards lightcone and thus are timelike separated from A.[29] Under the very plausible assumption that a violation of Bell inequalities implies a violation of locality, it seems to follow that QM is in conflict with STR, which in turn would leave the NST without any sufficient motivation. This, in fact, is the conclusion that Bell came to:

> For me then this is the real problem with quantum theory: the apparently essential conflict between any sharp formulation and fundamental relativity. That is to say, we have an apparent incompatibility, at the deepest level, between the two fundamental pillars of contemporary theory [...] (Bell 1987, p. 172)

To see whether this conclusion indeed cannot be avoided, I wish to first point out, following Seevinck (2010), that the NST should not be regarded as *proving* that there

[28] For the technical details see Redhead (1987, p. 116).

[29] Healey (2014, sec. 2) argues, that in this way, which can be made mathematically precise, 'locality' should be understood, and not as a vague causality condition as it can be found in the writings of Bell.

cannot be any signalling with superluminal speed. Otherwise it would be begging the question against interpretations of entanglement that conclude exactly the opposite, namely that EPR experiments show that there is superluminal signalling. Instead, the NST is a postulate, standing at the beginning of relativistic QM, and as any postulate it can turn out to be false. Second, a violation of locality does not imply that there is any superluminal signalling that can be exploited for sending messages (Maudlin 2002, p. 84). Thus, if the NST is interpreted as merely precluding the possibility of a so-called 'Bell telephone', then it is not threatened by a violation of locality.[30]

However, the conflict seems to be deeper, in that STR is often interpreted as forbidding any superluminal signalling, whether a possible tool for communication or not. Assuming that the violation of Bell inequalities shows that there is superluminal signalling, even though it cannot be used for communication, does this show that STR is false? Seevinck (2010) believes that it does, but at the same time he lists nine possibilities of how both can be reconciled and live in 'peaceful coexistence'.

I think a peaceful coexistence along the following lines is most promising at the moment. As Brown (2007) shows, a fair case can be made that STR is only a meta-theory, which on its own tells us nothing about the world. STR merely puts one constraint on other physical theories and that is that they have to be covariant, i.e., transform under the Lorentz or the Poincaré group. If this is the case, then prima facie there is no conflict, because QM can be given a covariant formulation.[31] Furthermore, Maudlin (2002) shows in detail that STR does not forbid superluminal signals per se, but only those that form loops and thus would lead to paradoxes. Maudlin (2002) highlights that accepting superluminal signals without loops, might only come at a price, as it is not entirely clear what could do the job of forbidding loops. Nonetheless, as it stands, STR can be given an interpretation in which a conflict with entanglement is avoided. Whether this minimal interpretation of STR is the right one remains debatable, of course.[32] In conclusion, I wish to assume for the following discussion of causation that QM can be given a relativistic formulation, motivating the NST, and that relativity is not threatened by entanglement, even if the locality condition is violated. This remains a contentious claim. However, the present work, being an exercise in interpreting QFT, relies in many ways on the validity of STR, and given the countless results that confirm it, I believe it to be reasonable to adhere to it as long as possible.[33]

[30]Some experimenters take it as a challenge to find out how fast a signal that connects two measurements in an EPR setting, violating locality and possibly STR, has to be. One of the latest results shows that if "a privileged reference frame exists [in which the signal travels] and is such that the Earth's speed in this frame is less than 10^{-3} times that of the speed of light, then the speed of the influence would have to exceed that of light by at least four orders of magnitude." (Salart et al. 2008, p. 861)

[31]It should be mentioned, though, that there are still some unresolved mathematical issues (cf. Jaffe 2007).

[32]See also Weinstein (2006) who concludes that STR does not forbid superluminal signals with a slightly different argument than Maudlin.

[33]I believe that this is also the stance taken by most physicists. For example, 't Hooft (2007,

Before turning to causation, I briefly wish to address the question whether entanglement and the strange EPR correlations are merely special phenomena, appearing only in carefully arranged experiments, or whether they are an ordinary thing in the world. So how do entangled states come about in the first place? If within the von Neumann formalism of QM the state $|\psi\rangle = \sum_n a_n |\psi_n\rangle$ interacts with the state $|\phi\rangle = \sum_m b_m |\phi_m\rangle$, in general the result will be the entangled state $|\Psi\rangle = \sum_{nm} c_{nm} |\psi_n\rangle |\phi_m\rangle$ with $c_{nm} \neq a_n + b_m$. It can be shown (cf. Gemmer & Mahler 2001; Durt 2004) that a separable state $|\Psi\rangle_{Pro} = \sum_n a_n |\psi_n\rangle \sum_m b_m |\phi_m\rangle$ only stays separable if it is evolved by a product operator $\hat{U} = \hat{U}_\psi \otimes \hat{U}_\phi$, where $\hat{U}_\psi$ only acts on $|\psi\rangle$ and $\hat{U}_\phi$ only acts on $|\phi\rangle$. Product operators, however, cannot describe interactions. From here one might infer that interactions in QM always result in the entanglement of the previously separable interacting states and this might lead to the somewhat extreme position that "if everything is in interaction with everything else, everything is generically entangled with everything else" (Bacciagaluppi 2012).

However, we have to be careful about what the theorem about product operators really tells us and what not. It is true that a state only remains separable if it is transformed by a product operator and that interactions cannot be described that way. On the other hand, it does not follow that all non-product operators and all interactions only lead to entangled states. It is crucial to realise that the production of entanglement by a unitary operator is a reversible process, as most processes in nature are. Taking the hermitian conjugate of an entanglement producing operator can transfer an entangled state into a separable one. It therefore seems reasonable to assume that there are interactions that produce entanglement as well as interactions that destroy them. There is no reason to assume that entanglement only works in one direction but not in the other. Additionally, combinations of these two processes might happen in a single interaction. Furthermore, Busch (2003) has shown that interactions can also be described by an operator that has the form of a "swap map". Such an operator will transform a separable state into another separable state and only lead to short lived entangled states in between. Having said that, Busch concludes: "Whether such measurement dynamics can be implemented by realistic interactions is another question." As in relation to QFT, scattering processes do (generally) not lead to the entanglement of the initial scattering particles. For example the result of a $e^- + e^+$ scattering is not an entangled state of $e^- + e^+$, contrary to what follows from the formalism of QM. Though it is reasonable to assume that quantum field theoretical interactions will lead to entanglement between certain degrees of freedom of the final state. However, it is hard to estimate what degrees of freedom are entangled to what extent. To answer this question we would at least need an explicit calculation for a simple realistic interaction. To my knowledge, no such calculation exists let alone any general theorem. In conclusion, entanglement is certainly something that can happen in very normal circumstances and not just

p. 661) brings it to the point: "Thus, in the first half of the twentieth century, the question was asked: '*How should one reconcile Quantum Mechanics with Einstein's theory of Special Relativity?*' [...] Quantum Field Theory is the answer to this question."

in experiments. However, it is unwarranted to conclude from here, as Bacciagaluppi does, that everything is entangled with everything.

6.4.2 Causation

The fact that measurement results in EPR experiments are perfectly correlated has lead many to believe that this is not a matter of mere correlation, but causation. Somehow the measurement results must be causally connected, or else it would come close to a miracle that they are always correlated. Thus, so one might believe, the question is not whether there is a causal connection, but only what the nature of this connection is, that is, what theory of causation it captures best. If this common belief was true, then it would pose a considerable obstacle for the present work, most obviously in the following way. The theory of causation I am hoping to defend is a theory in terms of the transfer of a conserved quantity. Given that locality is violated by entanglement and the probabilities for measurement results in EPR experiments are not screened off from another by the state and the apparatuses settings prior to the measurement, then they cannot be explained by a local common cause.[34] On the other hand, there is no known signal from one measurement to another that could be thought of as transferring causal influence. Given the plausibility of the assumption that entanglement is a causal relation, the obvious conclusion seems to be that there is a causal relation that cannot be described within a transfer theory and that therefore the latter is either limited or false. In this section I will argue against this conclusion. The basic line of thought will be that entanglement is not in conflict with transfer theories alone, but more generally that hardly any of the platitudes of causation from chapter 3 above is applicable. As a consequence, I conclude that under no theory of causation that shares these platitudes entanglement turns out to be a causal relation and that therefore the intuition to the contrary is a meaningless claim; if it is still maintained that entanglement is causal, then it is simply not clear what the term 'causal' is supposed to mean.

Counterfactuals Most commonly, entanglement is associated with counterfactual theories of causation. For example, according to Esfeld (2001, p. 175) "Bell experiments show a counterfactual dependence between space-like separated measurement outcomes [...], they thus satisfy Lewis' criterion for causation." And Maudlin (2002, p. 126) is content that, "[f]ortunately, we need not produce a completely general theory either of causation or of counterfactuals. The case before us involves the most uncontroversial application of these notions, examples on which all acceptable theories must agree in order to be acceptable." There are good reasons to believe

[34]Spelling out in detail what this means and proving that this statement is correct is far beyond the scope of the present work. The most precise discussion can be found in Hofer-Szabó et al. (2013, ch. 9). But see also Maudlin 2002, p. 125 and Fenton-Glynn & Kroedel (2015, p. 2). Also Rédei (2002) for a general discussion of Reichenbach's common cause principle, and Suárez (2004) for a survey of possible causal and in particular common cause explanations of EPR correlations.

that a theory of causation based on counterfactuals is not feasible (as I have argued in section 3.3.1 above), however, for the sake of the argument I will assume the contrary. The most important question to start with then is which counterfactuals hold in the case of entanglement.

It can be shown (cf. Butterfield 1992a; Healey 2014) that, if locality is violated, the following two counterfactuals are true:

(i) 'If Alice had measured A, then the probability for Bob to measure B would have been different.'

(ii) 'If Bob had measured B, then the probability for Alice to measure A would have been different.'

Is this enough for a causal connection between A and B? For Maudlin (2002, p. 127) it is, given that "[o]ne key way that causation goes beyond mere correlation, then, is that causal connections support counterfactuals while *de facto* correlations may not." Also, Butterfield (1992b,a) comes to the conclusion that the condition of outcome independence is equivalent to counterfactual causal independence, and that consequently from a violation of outcome independence it follows that there is superluminal causation. This, however, has been criticised by several authors (Skyrms 1984; Elby 1992; Fenton-Glynn & Kroedel 2015; Healey 2014), and I want to briefly present their objections.

It is obvious that the above counterfactuals (i) and (ii) are symmetric. This is a consequence of the possibility that the measurements of Alice and Bob are spacelike separated, and that therefore according to SRT they have no objective temporal order. The problem for the counterfactual analysis is now precisely that both, (i) and (ii), come out as true.[35] In a reference frame in which A comes before B the first counterfactual is true, hence A causes B, and in a reference frame in which B comes before A the second counterfactual is true, hence B causes A, and under a standard interpretation of STR, both reference frames are equally possible and justified. Skyrms (1984, p. 246) concludes from here that the measurement results, A and B, are "forming a rather odd, closed causal chain", in which each is the cause and the effect of the other. Fenton-Glynn & Kroedel (2015, sec 3.2), on the other hand, come to the slightly different conclusion that the counterfactual theory is "indecisive" about whether entanglement is a causal relation, since it gives no handle to decide between (i) and (ii). Be that as it may, both conclusions are equally problematic and show that a counterfactual theory of causation is not applicable to entanglement.[36]

[35] There are further issues here, as to how the truth values of the counterfactuals should be evaluated, which I will not discuss for the sake of brevity (cf. Fenton-Glynn & Kroedel 2015).

[36] The situation for standard non-probabilistic counterfactuals is even worse, in that in any reference frame counterfactuals of the form 'If measurement result A would not have happened, result B would not have happened' will come out as false. Given that, say, the unconditional probability for Bob to measure $|\uparrow\rangle_b$ is $1/2$, it is possible for him to measure $|\uparrow\rangle_b$, even if Alice did not measure $|\downarrow\rangle_a$.

Asymmetry On top of that, I wish to maintain that the missing temporal order between spacelike separated events is not only a problem for counterfactual theories, but for causal theories in general. As I have argued in section 3.3.5, the causal relation is temporally directed in that the cause comes before the effect and not vice versa. The problem is that the strategy for defining a temporal order between events that I followed in section 4.3 is not an option in the case of spacelike separated events, and I know of no alternative that would be. In conclusion, any causal relation between measurement outcomes in EPR experiments must be temporally symmetric.

Maudlin (2002) seems to accept this conclusion and only argues for a relation of 'causal implication' and not causation. The difference between both is that "we do not suppose that it follows from the fact that A is causally implicated with B that A caused B or B caused A." (Maudlin 2002, p. 128) This, however, seems to be a very unilluminating response that does not contribute to the debate, since it is simply a change of concepts. The question is not whether entanglement is some symmetric relation called 'causal implication', but whether it is a causal relation.

Interventions The next problem for many theories of causation in the case of entanglement comes when trying to apply some form of interventionism as a criterion for causation. Healey (2014) has identified three reasons why interventionism does not work for entangled states. Number one is that the supposed causal relata, the measurement outcomes, are not suitable as variables that can be intervened on:

> Since the values of the variables A, B [i.e., measurement outcomes] are represented as fixed elements in fine-grained quantum models, they are not exogenous, and so they are not free elements even in the broad sense implicit in the theoretician's use of quantum theory. Nor are they subject to any physically possible intervention. [...] Quantum theory itself provides no resources on which one can draw to make sense of an intervention capable of changing the outcome of Alice's or Bob's measurement [...]. (Healey 2014, p. 8)

Accordingly, any form of interventionism that relies on whether manipulating one variable will change another is not applicable.

Problem number two comes from STR again. Having established the physical impossibility of interventions as a result of the first problem, Healey now considers their conceptual possibility. According to Woodward (2003), if there is a causal connection between A and B, then interventions on A have to switch-off any previous causal influence on A, such that the value of A is completely determined by the intervention. Now, if both counterfactuals, (i) and (ii) are true, then it follows that we can intervene equally on A or B to manipulate the other. In other words, "it makes sense to intervene on B if and only if it makes sense to intervene on A—a natural assumption in the light of Lorentz invariance." (Healey 2014, p. 8) It follows that any proper intervention, in Woodward's sense, on A has to switch off B as a cause of A, and vice versa, since otherwise the intervention would not be the only cause of A (or B). Hence, if counterfactual (i) is true, then (ii) must be false, and vice versa. Since both counterfactuals are equally justified, it follows that if they are true, then they are false. This reductio shows that interventions on entangled states are not even conceptually possible. It might be objected that the above arguments rest on the spacelike separation of the measurements and therefore do not hold when they are timelike separated. However, Healey (2014, p. 9) notes that there is no reason to think that a causal connection for timelike separated events can be established that then can be extended onto the spacelike case, and it seems reasonable to ask for a theory of causation that works for both.

The third argument against interventionism in entanglement comes from no-signalling. Even if the measurement results A and B were causally connected, it is a consequence of the NST that it would be impossible to measure any change in B as a result of an intervention on A, and vice versa. Since the statistics for Alice do not change, no matter what Bob does, it is impossible for Alice to notice any intervention done by Bob. This might not be a strong argument against causation per se, given that there might be causation even if we cannot detect it, but at least it makes it hard to see how interventionism could be applied (cf. Skyrms 1984, p. 246; Healey 2014, p. 10)[37]

Relata Having argued for the impossibility to apply a counterfactual theories of causation, the criteria of asymmetry and interventionism, I now wish to point out that there is also a difficulty in defining what the relata of a supposed causal relation should be. Maudlin (2002, p. 8) starts from the assumption that an entangled state is composed out of "separated pairs of systems"; an assumption that he never justifies, and that I want to call into question. The basic problem comes from the fact that an entangled state is not constituted out of two or more separable states. To be more precise, based on Einstein (1948), Howard (1989, p. 225 f.) formulates the following 'separability principle':

> It asserts that the contents of any two regions of space-time separated by a nonvanishing spatio-temporal interval constitute separable physical systems, in the sense that (1) each possesses its own, distinct physical state, and (2) the joint state of the two systems is wholly determined by these separate states.[38]

As Howard explains, however, taking spatiotemporal distance as a criterion for separability can lead to the unintuitive consequence that every space-time point of a field constitutes a distinct infinitely small system. Furthermore, Esfeld (2001, p. 183) points out that there are cases of entangled states in which the supposedly entangled entities are not spatiotemporally separated. Generally, it is not possible to assign (non-temporal) parts of one entangled state to different space-time points. Hence, to discuss the question whether an entangled state can be regarded as composed out of distinct entities, which then could be the relata of a causal relation, it seems

[37] A further argument against interventionism could be based on the 'robustness' criterion that Redhead (1987) introduces. However, see Maudlin (2002, p. 150 ff.) for a critique.

[38] At this point, it is sufficient to discuss merely qualitative notions of separability. For more precise quantitative notions, see my presentation in section 6.3.3.

reasonable to drop the requirement of spatiotemporal distance. Accordingly, I will assume the following separability principle:

> *State Separability*: The state assigned to a compound physical system at any time is supervenient on the states then assigned to its component subsystems. (Healey 2009)

Howard calls his separability principle a "fundamental ontological principle", but I believe that at least in the case of state separability one does not have to go that far to motivate its validity. Rather, state separability seems to reflect a basic way how physics works. A state is a certain set of physical properties, for example energy, charge and spin. A certain set of these properties describes a fundamental entity, if according to the theory it cannot be divided further. This classification in fundamental entities is of course dependent on the theory and if the theory changes, the classification may change, e.g., when new smaller particles are found. Compound systems, on the other hand, can be described as many particle systems, that is, by describing multiple fundamental entities in one state. In that case the state of the compound system is fully determined by its constituents. This is not supposed to mean that composition always goes by way of a simple sum of the parts. There might be multiple and diverse composition relations. However, 'state separability' is committed to the view that a composed system is decomposable into its parts.

As is well known, an entangled state is non-separable and therefore cannot be regarded as a compound system according to 'state separability', that is, an entangled state cannot be factorised into two or more pure states.[39] To add a minor, maybe only terminological point, I do not agree with authors (e.g. Esfeld 2001; Healey 2009, sec. 8) who say that entanglement *violates* separability. Given that physics does not allow one to decompose an entangled state into sub-states, which would have properties like spatiotemporal location and energy-momentum, one can also conclude that an entangled state is not a compound physical system, rather than that it is a non-separable compound physical system. The latter conclusion indeed would violate 'state separability', while the former does not. Having said that, there are interpretations of entanglement, such as the holism discussed below, of which reasonably can be said they lead to a violation of separability, even though according to them an entangled state is not composed of other states.

In the light of non-separability, what could be the relata of a causal relation (remembering from section 3.3.1 that causation needs distinct relata)? It cannot be the component states of an entangled state, because there are none (cf. Skyrms 1984, p. 246). Another possibility would be given, according to Esfeld (2001, p. 181), "[i]f we focus on the changes in the states of the two measuring instruments which register the outcomes, we clearly have two distinct events, and these events are correlated." However, as argued above, the measurement outcomes are not suitable relata, since they are not variables that can be intervened on.

Some philosophers (e.g. Teller 1986; Howard 1989; Esfeld 2004) interpret non-separability in the way that an entangled state forms one holistic system. Healey

[39] Again, see my discussion of separability in section 6.3.3 above, also Healey (2009, sec. 5).

defines holism as follows:

Holism: “As applied to physics, ontological holism is the thesis that there are physical objects that are not wholly composed of basic physical parts.” (Healey 2009)

An immediate conclusion from here is that a radical holistic interpretation in which an entangled state has no parts at all is incompatible with causation, simply because there are no parts that could be causally connected.

However, there are less radical holistic interpretations. For example, Esfeld (2004, p. 608 f.) comes to the conclusion that an entangled state is a holistic system in the following way: “In any case of quantum entanglement, only the joint state of the whole completely determines the probability distributions of the state-dependent properties of the parts by determining correlations among these probability distributions.” I see two problems with this statement in the light of a causal interpretation. First, by noting that an entangled state has probability distributions as parts, Esfeld, to my mind, refers to the fact that an entangled state can be decomposed into mixed states described by von Neumann density matrices, which, however, do not uniquely determine the entangled state (cf. Healey 2009, sec. 5). Nonetheless, it is not obvious that von Neumann density matrices can equally well be interpreted as describing physical entities as, say, rays in Hilbert space, since common properties of physical states cannot be attributed to them, as also Esfeld (2004, p. 610 f.) notes. Be that as it may, the second problem is that, as Esfeld acknowledges, only the whole entangled state determines the correlations between the measurements. I do not see how that could be reconciled with the idea that there are causal connections between the parts of an entangled state that are responsible for the correlations. In other words, if, according to the definition of holism, the whole entangled system has properties that do not belong to any of its parts and these properties are responsible for the correlations, then the correlations do not come about by one part causing the other. Furthermore, given that the most plausible explanation for EPR correlations is a violation of locality, the measurement results are not screened off from another and the whole entangled state cannot be the common cause of the measurement outcomes.

All these worries notwithstanding, one might reply that whatever is needed to make the whole entangled state is not something extra to the parts. How this is possible becomes clear in the ‘relational holism’ advocated by Teller (1986, p. 73), which is “the claim that objects which in at least some circumstances we can identify as separate individuals have inherent relations, that is, relations which do not supervene on the non-relational properties of the distinct individuals.” As Teller argues, since in the case of entanglement we cannot assign any definite value of the entangled property to any entangled particle, we should regard the entangled property as a non-supervenient relation. The sense of supervenience that entanglement arguably supports is further clarified in French (1989). He shows that entanglement as a relation can be regarded as strongly non-supervenient in the following sense:

> *Strong Non-Supervenience*: a relation is strongly non-supervenient upon a determinable non-relational attribute if the appearance of this relation is

> neither dependent upon nor determined by its relata bearing the non-relation concerned. (French 1989, p. 10)

On this basis, Esfeld (2004) argues for a relational holism in which the entangled systems have no other properties other than the non-supervenient entanglement relations they stand in. As a consequence, "[t]here are no qualitative properties whatsoever—not even relational conditional probabilities—that distinguish one such system from all the other ones." (Esfeld 2004, p. 612)

The above exposition of holism necessarily had to omit many details. However, as I wish to stress, the goal at this point is not to discuss whether holism is a feasible interpretation of entanglement, but only whether it supports a causal interpretation; and to evaluate this, I believe, we have gathered enough material. In section 3.3.1 above, I have argued that the relata of a causal relation have to be distinct. Furthermore, I have analysed 'distinctness' in the concepts of 'distinguishability' and 'ontological independence'. I believe, that the holism as sketched above renders both of the latter concepts inapplicable. First of all, any two or more entangled systems ontologically depend on each other in the sense that one necessarily only exist as an entangled system, if another system, it is entangled with, exist. This follows from two premises. First, that any entangled system is fully constituted from non-supervenient relations. Second, that there is no relation without relata, which at least Esfeld (2004, p. 603) explicitly endorses. Hence, any entangled system can necessarily only exist as entangled system, if there is another system it is entangled with. With respect to distinguishability, I can only repeat what has already be mentioned several times, namely that the relation of entanglement is not sufficient to distinguish one system from the other (cf. French 1989, p. 11). In conclusion, even if an interpretation of entanglement is adopted in which an entangled state is a holistic system of non-supervenient relations, I see no grounds to identify the parts of that state as sufficiently distinct to be the relata of a supposed causal relation.

To conclude, the main points of this section were that a supposed causal relation of entanglement cannot be captured in a counterfactual theory, is not temporally asymmetric and does not fit with interventionism. Furthermore, measurement results are not the right kind of relata for a causal relation, since they cannot be intervened on, and if the entangled state is considered before the measurement, then it seems impossible to find any relata for the causal relation. It cannot be denied that entanglement has the strong 'feel' of a causal relation with it. A failure of locality implies that one measurement can alter the probability for another measurement result, and therefore one might feel like there must be a causal connection between both. However, as the history of the debates around causation has shown again and again, very special conditions have to be fulfilled for a correlation to be a causal relation. Entanglement does not give support for these conditions and therefore is not a causal relation under any theory that shares the most common intuitions about causation. Hence, the fact that entanglement cannot be described as a causal relation within a transfer theory of causation is not a defect of the latter, but follows from the non-causal nature of entanglement.[40]

6.5 Conclusion

In this chapter, I have argued that the most threatening problems from physics for a causal interpretation of QFT can be answered to a satisfying level. First, the proof of Haag's theorem is made impossible by the very same methods that are today's standard in QFT. Hence, the interaction picture does exist and QFT can be interpreted as describing interactions between fields. In the case of the measurement problem, I have argued that it would be hastily to conclude that the problem is unsolvable by physics alone. Instead, I have tried to motivate a more serious engagement with the question whether QFT can help to understand the physics of measurements. This result is not as clear cut as might be desirable. Nevertheless, the prospects for the causal interpretation are arguably better than before. Finally, I have argued that hardly any of the widely shared intuitions about causation is apt for the case of entanglement. In conclusion, entanglement is not a problem for transference theories of causation, rather it should not be regarded as a causal relation tout court.

There are certainly topics left that I could not treat in depth here. For example, it would be interesting to spell out the consequences for the causal interpretation that possible solutions of the measurement problem might have that differ from standard quantum theory. Also, it would be very helpful to know more about the role of entanglement in QFT, or to analyse modern methods of manipulating entangled states, such as entanglement swapping. These topics, unfortunately, must be left for a future investigation elsewhere.

[40]In this discussion I have assumed that a local common cause that screens off the measurement results from one another can be out ruled and that the supposed causal influence is a direct cause from Bob's measurement result or detector setting to Alice's result or setting (or vice versa). However, Näger (2013) has shown that the situation is slightly more complicated, in that there has to be at least one non-local common cause, or in other words, a violation of outcome independence is not sufficient to violate the Bell inequalitites. Näger argues that in order to violate the Bell inequalities, the entangled state $|\psi\rangle_{Sin} = {}^1\!/\!\sqrt{2}\left(|\downarrow\rangle_a |\uparrow\rangle_b - |\uparrow\rangle_a |\downarrow\rangle_b\right)$ must first go through one detector where it collapses into the product state $|\psi\rangle_{Pro} = {}^1\!/\!\sqrt{2}\left(|\downarrow\rangle_a |\uparrow\rangle_a - |\uparrow\rangle_b |\downarrow\rangle_b\right)$, and second that this product state is the common cause of both Alice's and Bob's measurement outcomes that does not screen off. This model has the advantage that it makes clear what the relata of the supposed causal relation are, namely one detector setting as the cause of both measurement outcomes. However, with respect to a possible causal interpretation, I whish to point out that there are problems both with asymmetry and interventions. First, Näger (2013, p. 37) presupposes that one measurement comes before the other in time. Clearly, in a relativistic theory, this precedence does not exist and one runs into the problem that I discussed in the 'Asymmetry' paragraph. Second, even though one can intervene on and controle detector settings, the NST prohibits that one can infer the detector setting from the mesurement outcome. Besides these problems, it is currently unclear to me whether this non-screening-off common cause structure can be spelled out without introducing an additional direct causal relation between the measurement outcomes, which would run into the same problems that I discussed in this section.

7 Synthesis: A new transference theory of causation

In this chapter, I wish to combine the results of previous chapters, in order to create a new version of the CQT. According to the methodology presented in section 3.2, the first step in finding a theory of causation consists in analysing the concept ‘causation’ into widely shared intuitions. Then, in the second step, it has to be investigated in how far these intuitions correspond to anything in physics. Accordingly, in sections 3.3.1 to 3.3.6 I have discussed six intuitions about causation, and in chapters 4 and 5 I have taken a closer look into physics in order to circle out features that may or may not serve as realisers. It is now time to take the second step and answer whether those features in fact can realise the intuitions. Consequently, in this chapter I will, one after the other, go through the basic intuitions about causation and ask for each whether they are realised by QFT. Doing so, I will use Dowe’s CQT as a background and guiding theory that I will retain when possible, and alter when necessary. Finally, I will summarise the results into a new theory of causation for QFT.

I wish to point out again that this grounding of the concept ‘causation’ in physics is not some kind of necessary identity. I completely admit that it can coherently be denied that the features of physics I rely on are anything like what could be called causal. Rather, this grounding is a suggestion for how problematic metaphysical terms can be understood. At the same time, I agree with Chakravartty (2013) that the grounding relation is potentially problematic and needs to be spelled out in further detail. However, at the moment, taking it for granted is the best the naturalised metaphysician can do.

Relata, group structure and local conservation laws In section 3.3.1, I have argued that one of the most basic platitudes about causation is that it is a relation between distinct relata. Thus, the obvious first question is whether there is anything in QFT that could be regarded as distinct relata. As I wish to suggest, the group theoretical structure of QFT together with the dynamics as constrained by the Lagrangian are sufficient to define these relata up to the point necessary for a theory of causation. The role of group theory has been made precise in section 5.3.2. We found that the kinds of particles that play a role in QFT are distinguished by the group representations under which they transform. Additionally, from the transformation property of the Lagrangian of the theory we know which kinds of particles play a role in it and from the initial and final states of a specific process, given, e.g., in the S-matrix, it is determined which particles appear in each state respectively. Consequently, whatever takes part in an interaction of QFT has to be an irreducible representation of one of the relevant symmetry groups, singled out by the specific values of rest-mass and spin, or, if the interacting object is not an elementary particle, but, e.g., a molecule, consist of irreducible representations. It is not group structure alone, that provides this information, but a particular representation of group structure, given by the Lagrangian.

On top of that, initial and final states in QFT fulfil the requirements that causal relata are distinguishable and existentially independent. First, they are distinguishable by their places in spacetime. This is already ensured by the condition that the initial state exists in the past and the final state in the future of an interaction, but usually these states will also be located at different positions in space. Second, they are existentially independent in that always exclusively either the initial or the final state of a single interaction exists, but not both at the same time. Hence, the group structure of QFT with the Lagrangian give the answer on the question how to identify distinct relata of a putatively causal interaction, namely, as the initial and final states as defined by group theory.

In order to make the recourse to group structure more convincing, I wish to argue that group structure together with local conservation laws, which are determined by the symmetries of the Lagrangian, can solve a problem that Dowe sees for transference theories of causation. That is the problem of spacewise and timewise gerrymanders. In general, a gerrymandered object is something that might be regarded as the relata of causation, but for which there are good reasons to exclude it from that role. A timewise gerrymander would be, for example, an object that is my desk before time t_1 and my computer afterwards. A spacewise gerrymander on the other hand could be an object that consists of my desk and my computer on it at every time. The famous spot of light moving on a wall is a timewise gerrymander, since it consists of different photons at every point in time. Both kinds of objects have to be excluded from causal relations, since they lead to the certainly unintuitive result that, e.g., every time my computer enters a causal relation my desk does as well, and vice versa.

Dowe relies on ontological commitments to rule out gerrymandered objects. Again, Dowe takes an object to be "anything found in the ontology of science (such as particles, waves and fields), or common sense (such as chairs, buildings and people)." (Dowe 2000, p. 91) This immediately rules out spacewise gerrymander, since they do not belong into any sensible ontology. Furthermore, timewise gerrymanders are ruled out by the criterion that objects possess an identity over time (cf. Dowe 2000, p. 100). So while a photon counts as an object, several photons appearing at different points at different times do not possess a single identity. In what follows I will argue that the role can be played by conservation laws in the form of continuity equations and group structure. The former rule out timewise gerrymanders and the latter spacewise gerrymanders.

In section 5.3.3, I have presented Noether's theorem, which connects every continuous symmetry of a physical system with a conserved quantity that system has. More precisely, the resulting conserved quantities are locally conserved, which means the following. It is in accordance with *global* conservation laws that a conserved quantity vanishes at one point in space and simultaneously appears at another distant point. As long as the amount of the conserved quantity remains the same, the global conservation law is satisfied. On the contrary, such a process is forbidden by *local* conservation laws. This requirement can be expressed by continuity equations, which are conservation laws put into the form of a differential equation. The essence of a continuity equation is that either the amount of a conserved quantity is invariant in a certain region of spacetime or there is a current of the quantity though the surface of the region. Consequently, if a locally conserved quantity vanishes at one point in space and appears at another point later, then there must have been a current of that quantity between the two points (cf. Greiner & Reinhardt 1996, ch. 2.4). If we now take a closer look at timewise gerrymanders, it becomes apparent how continuity equations help. Timewise gerrymanders consist of different objects at different times and even though they might have the exact same properties, like photons of a spotlight which is moving on a wall, they are not governed by a continuity equation. In particular, the continuity equation excludes that different temporal stages of a single, say, photon coexist, while this is perfectly possible for different photons emitted by a spotlight at different times. This shows that a photon might be a causal process while a spot of light certainly is not.

However, a worry that one might have is that, contrary to the identity of objects, conservation laws only imply numerical invariance. As a consequence, it seems to be in principle possible that, say, photon a is at one point in time replaced by another photon b, which is exactly like the first one. If a and b are not distinguished by their respective identity, both photons together would form a timewise gerrymander. I believe that identity is unnecessary to argue against such cases. First of all, the example is question begging in that it already presupposes that the photons have an identity, and that it makes a difference whether the process consists of photon a or b. Now, this question is one for which an answer cannot be found within physics, since physics does not describe the identity of objects. Therefore, I suggest that from a physicalist perspective there is no fact of the matter whether photon a is replaced by b or not, as long as no laws of nature are violated. As far as I am concerned this is a pseudo problem, constructed by the unjustified introduction of identities.[1]

For Dowe, identities also solve the problem of spacewise gerrymanders. What is it that makes us differentiate between two particles in the first place? Why not take them as a single one at any give point in time? Following Dowe, it is again the ontology of science and the identity of objects. If according to the ontology of physics photons are fundamental particles then it obviously makes no sense to treat two of them as if they were a single one. However, again physics gives us the tools at hand to distinguish fundamental particles without recourse on ontology. From section 5.3.2 it is clear that elementary particles can be described as irreducible representations

[1]Though, this point surely depends on which criteria one takes to define identity. There might be theories of identity according to which the numerical invariance following from continuity equations is sufficient for identity.

of certain symmetry groups, and kinds of particles can be distinguished by their particular rest-mass and spin. This together does not suffice to make a difference between particles of the same kind, i.e., to give each particle an identity, however single particles can be distinguished from composite systems by the amount of the characteristic properties they have. So, for example, a composite system of two electrons will have a higher rest-mass than a single electron. Now, since spacewise gerrymanders are aggregates of objects, they can be distinguished from fundamental objects by the means just described.

As a final remark, I wish to point out that by defining the relata in this way, no further decision on an ontology for QFT is made. There is a large, and as of today, inconclusive discussion about what kind of entities exist according to QFT, that is, particles, fields or something else (cf. Kuhlmann et al. 2002). A theory of causation can remain neutral with respect to ontology, as what matters is only that there are distinct relata, and in general particles can be the distinct relata just as well as fields or what not. I see it as an advantage that with the help of the concept 'causation' progress in understanding QFT can be made, without having to dive into the deep and dangerous waters of ontology.

Intersections and locality The next questions are, whether initial and final states form world lines and whether they intersect; pointing to group theory by itself does not give any answer to that. If these questions can be answered affirmative, then world lines can still play the role as defining the place where interactions occur, as they do in Dowe's CQT. However, one of the results of section 5.2 was that there are considerable difficulties in identifying world lines in QFT. Initially it seemed promising that even in field theories, with fields defined over all of spacetime, world lines that do not fill out the whole of a Minkowski diagram can be defined by taking only the region of spacetime into account where the field has values higher than a certain threshold. This proposal then was fleshed out following the work of Wallace by regarding a quantum field to be localised in a region where it significantly differs from the vacuum state. As we have seen, this definition of localised fields also agrees with experimental practice. In scattering experiments fields are localised by the use of collimators in a very small region, and tracks in cloud chambers, even though they are not world lines but a series of interactions, show that what is interacting has to be localised in a small region.

This might sound as if Dowe's world lines could be carried over into QFT, but we also came to the negative result that even though localised fields can be mathematically defined and have their applications in experiments, their general significance remained somewhat unclear. There is certainly no constraint in QFT for fields to be always localised in a small region. However, given that the role of world lines in Dowe's CQT is to specify the location of interactions, I wish to suggest that QFT has the means to replace them entirely. In section 5.3.4 we have seen that one of the most fundamental characteristics of QFT is that it is a local theory. In contrast to the approximate localisation discussed in the previous paragraph, the principle of locality is a general constraint in QFT. It results from the requirement of Lorentz invariance, and ensures that interactions have to happen at points in spacetime and cannot take place over spacelike distances. On top of that, from the spacetime parameters at which the interaction is evaluated, as given in the S-matrix, we know where interactions take place. Thus, the principle of locality teaches us that interaction in QFT goes hand in hand with intersection, provided that intersection is synonymous for taking place at points of spacetime, as opposed to action at a distance. I believe that this is a viable understanding of 'intersection', and it provides exactly what Dowe used his notion of intersection for. In consequence, Dowe's reliance of intersecting world lines can be replaced in QFT by the principle of locality.

The causal relation and forces Above, I have argued that the difficulties Dowe's CQT encounters in the context of QFT can be redeemed. On top of that, as I now wish to explain, there are things QFT teaches us that exceed what has been proposed by Dowe, and that may serve to draw the physical basis of causation even sharper. Most importantly, QFT gives more insight into what the physical relation is that constitutes causation than Dowe's CQT.

In section 3.3.1, I have argued in favour of the view of causation as an intrinsic relation, viz., as supervening on the natural properties of the relata and the natural relations holding between them. But what exactly is the causal relation in a physical world? For Dowe it is the exchange of a conserved quantity, but he fails to give any explication of how this exchange happens, other than that world lines intersect. In QFT, on the other hand, we can say more than that. From section 5.3.5 we know that interactions in QFT are not simply the meeting of whatever interacts, e.g., like a layman's picture of colliding billiard balls in which only the balls are involved. To the contrary, in QFT interactions are always mediated by a force. Furthermore, with the help of group theory forces can be categorised in the same way as other fields.

Thus, even though the conclusion about the existence of virtual particles was negative, I stressed that this does not imply that force fields do not play the role of intermediate fields in interactions. Forces in QFT do not work like action at a distance, but act in between initial and final states and have properties like energy or helicity. For causation this means that the force field can be regarded as an intrinsic relation, connecting initial and final state and supervening on them. First, the force field only exists as an intermediate state between initial and final state, and second, its kind and properties are completely dependent on the initial and final states, in the sense that they are determined by the initial and final states. Thus, the initial state will determine whether, say, only the electromagnetic force or also other forces take part in the interaction. I have to concede, however, that the meaning of the 'in betweenness' of forces cannot be made as clear as might be desirable. As far as I can see, QFT does not support an interpretation of 'in between' in a spatiotemporal sense. What we know for sure is that forces exist in between in the sense that initial and final states are free and do not interact with each other. I would suggest, this is all we can say. It might simply be a case of where QFT does not license an intuitive picture, 'picture' in a literal sense, which Dowe's CQT draws, namely that of a causal process as a sequence of connected world lines. Still, I want to maintain that QFT is a step forward in clarity.

Conserved quantities and interventions The next advantage over Dowe's CQT is that in QFT we can exactly pin down which conserved quantity has to be transferred for a process to be causal; a point that has been left rather vague before, as Dowe speaks only of conserved quantities in general. Given the plethora of conserved quantities in physics, some of them always conserved, some only conserved under special conditions, one wishes to know more than what Dowe provides (cf. Lupher 2009). In contrast, for the following reason I wish to suggest that the only quantity that is exchanged in a causal interaction is energy.

In section 5.3.3, we found that energy is locally conserved and present in all interactions in QFT. Already these two particularities, locality and a certain 'omnipresence', make energy an obvious candidate for what is exchanged. On top of that, there is a further reason from interventionism why, above any other property, it has to be energy. In section 3.3.4, I have argued that a theory of causation should comply with the intuition that the effect can be manipulated by manipulating the cause. This intuition has further been made precise with Woodward's (2003) theory of interventionism. For Woodward, causation is a relation between variables, which are properties that can take on more than one value. This is a necessary condition for interventionism, since the possibility to change the value of a variable is part of the meaning of 'intervention'. Asking now which property that is always conserved in interactions is a variable that can be manipulated, the answer is: only energy.[2] In experiments the energy of the initial state particles is precisely controlled, can be changed, and every change in the initial state's energy changes the energy of the final state.

To the contrary, another property that is always conserved is electric charge, but it is not a variable. Given the group structural ordering of particles, changing the charge of a particle amounts to changing the particle. On top of that, manipulating charge can practically not be done the way that energy can be manipulated. While we can change the energy of a particle, e.g., by accelerating it in a magnetic field, a particle with a different charge has to be created through a scattering or decay process. The problem with that is that creating a new particle that subsequently interacts cannot be counted as manipulating a causal process, but rather is the replacement of one causal process with another. Hence, if we want to find out whether the electrons we have produced are causally related with the electrons we have measured, it is not helpful to replace the electrons with muons. Rather, we will change the energy of the electrons in the initial state and see whether the energy of

[2] This is not to say that the application of Woodward's interventionism to physics is entirely unproblematic. However, due to restricted space I cannot discuss this matter here. A good defence of interventionism in physics can be found in Frisch (2014, ch. 4).

the final state corresponds to the change.

If energy is the property that is transmitted in causal interactions, which role do other properties play? This issue deserves more space than I can give it at this point, since I believe that it has to be embedded into a broader account of what laws are. Nonetheless, in short my suggestion is that properties other than energy, such as charge or spin, play a regulative role in that they determine which causal processes are possible and how they happen. For example, electromagnetic charge determines the strength of the electromagnetic force. Hence, it determines how much energy is transferred by that force. If the force was stronger, more energy would be transferred. My contention on this matter is that, following French (2014, ch. 10), properties must not be understood as belonging to individual objects, but as dependent and ultimately reducible to features of structures. It has been argued by Psillos (2006) that causation necessarily needs non-structural properties, however, again I wish to side with French (2014, p. 216), who argues that "the conclusion that causal claims require objects and properties as truth-makers seems unwarranted." It is clear that this paragraph merely points into the direction of where I believe a viable stance on properties in causal structures resides, and would need a lengthy discussion that I cannot give here.

Probabilities To proceed to the next intuition, in section 3.3.3 I have argued that it can be expected from a theory of causation to support the intuition that the cause raises the probability for the effect, without reducing causation to a probabilistic relation. Accordingly, section 5.3.7 has shown that QFT is a probabilistic theory, viz., an initial state leads to a final state only with a certain probability. At the same time, the occurrence of an initial state always trivially raises the probability of the occurrence of its final state, compared to the probability in the case of the absence of the initial state. It follows that the account of causation developed here has to be probabilistic, and thus an event does not necessarily leads to another, but only with a certain probability. Furthermore, if initial states raise the probability for final states, then causes raise the probability of their effects.

Directionality Finally, in section 3.3.5 I have presented the causal relation as directed in two ways, namely, temporal and causal. It is a widespread view that neither direction has a basis in physics. However, contrary to that, in chapter 4 I have argued that the temporal direction of causation can be reduced to the direction of time, as given by the direction of local energy flux, while the causal direction can be reduced to asymmetric self-contained causal structures, spelled out in terms of inter-model dependencies.

A new transference theory of causation Briefly summarised, the most central claims of this section give a condensed view of how the platitudes (P1), (P3), (P4) and (P5) from section 3.4 can be reduced onto physics:

Relata and relation: The relata are defined via group structure, and given in initial and final states. The relation is given by a force that transfers energy. Both are further constrained by local conservation laws.

Probability: The probability for a cause to bring about the effect is given by the S-matrix.

Intervention: We can manipulate the energy of a final state by manipulating the energy of the initial state.

Direction: The directionality of causation, temporal and causal, is represented by local energy flux and dependencies in causal structures.

Now we have all the ingredients together for a new version of the CQT, that pays tribute to the physics of QFT. In the most scarce terms, we can define causation as follows:

$\mathbf{CQT_{QFT}}$: A causal process is a quantum field theoretical interaction, i.e., the transfer of energy from an initial to a final state via a force.

Any process in QFT can serve as an example for a causal process, such as the scattering of an electron and a positron as a cause that, transmitted by the electroweak force, has the effect of a creation of a muon anti-muon pair. Also decay processes can be incorporated, such as the beta decay, in which the cause is the decay of a nucleus and by transmission of the weak force, e.g., a neutrino positron pair is created.

The main argument in support of $\mathrm{CQT_{QFT}}$ is that the physics it circles out is in agreement with our intuitions about causation. Accordingly, within a methodology that starts with a conceptual analysis and then proceeds to identify a realiser in the world for the concept, $\mathrm{CQT_{QFT}}$ is a viable theory of causation. Following this method also means that $\mathrm{CQT_{QFT}}$ is a reductive theory of causation, that is, it defines causation in physical terms and thus reduces causation to the physical. This does not mean, however, that an eliminativism about causation follows. Quite the opposite, as Lewis (1970, p. 427) observes, "to define [theoretical terms] is to show that there is no good reason to want to do without them." In other words, it shows that the use of the term 'causation' when describing physical processes is well founded.

One objection at this point might be that the reduction is a trivial reduction by definition. By way of definition one can reduce any domain A onto a domain B simply by defining terms in A via terms in B. The reduction does not rest on specific properties of QFT and could be made for any physical theory of interactions, thus it is hardly illuminating, or so one might argue. However, I believe that this objection is not tenable. $\mathrm{CQT_{QFT}}$ primarily makes use of group theory to specify the relata of causation, locality to specify the points in spacetime where a causal process takes place, conservation of energy and forces transmitting energy to specify the relation. These characteristics are not necessarily true for all physical theories. For example, it could be possible that according to our best physics, the whole world is just one big entangled state in which no distinction between different relata of a causal relation would be possible. Or it could be true that energy is not a locally conserved quantity, and thus could come into existence out of nothing. In both cases $\mathrm{CQT}_{\mathrm{QFT}}$ would not be a viable theory of causation, and indeed it would be questionable if there were any causation at all. Thus it is not trivial that processes in QFT can be described as causal. The objection is misguided, because the method followed here is not a simple reduction by definition, but an identification of a concept with parts of the world.[3]

On the other hand, $\mathrm{CQT}_{\mathrm{QFT}}$ is not linked to QFT as rigidly as it first might appear, in that it does not rest on a specific formulation of QFT. This stands in contrast to, e.g., an interpretation of QFT such as presented in Teller (1995) that is entirely based on the canonical quantisation method with its operator fields. It is unclear how Teller's account can deal with the completely equivalent, but differently formulated, path integral formulation of QFT. $\mathrm{CQT}_{\mathrm{QFT}}$, on the other hand, does not have such restrictions. All the properties of QFT on which $\mathrm{CQT}_{\mathrm{QFT}}$ rests are properties of the canonical quantisation as well as the path integral method, and I cannot find any features of the path integral method that would oppose $\mathrm{CQT}_{\mathrm{QFT}}$. In conclusion, I see this generality as a major advantage of $\mathrm{CQT}_{\mathrm{QFT}}$ over other interpretations of QFT.

[3]See Heathcote (1989, p. 102) for a further discussion of this objection.

8 Conclusion

In this dissertation, I have argued that QFT can be interpreted as describing causal processes. I have analysed the concept 'causation' into different widely accepted intuitions, and shown for each how they can be reduced to a feature of QFT. In particular, I have argued in favour of a novel way of how the directionality of causation can be reconciled with the symmetry of physics. Furthermore, I argued that group structure, local conservation laws, locality and forces can serve as the basis for a theory of causation. The result was that causation in QFT can be understood as the transfer of energy from an initial to a final state via a force.

It is worth pointing out again that this reduction must not be understood as an elimination of causation; rather the opposite. $\mathrm{CQT}_{\mathrm{QFT}}$ has the advantage of being explicitly defined in physical terms. Thus, under the assumption that it is sufficiently clear what the physical terms mean, we have a clear definition of causation. It has often been criticised that causation is an opaque concept that therefore should better be abolished. Given the reduction to physical terms, however, this is not the case. $\mathrm{CQT}_{\mathrm{QFT}}$ provides a plain and intelligible theory of causation, and these attributes can serve as a justification for further use of the concept.

This is a significant result, considering that causation today is a ubiquitous concept across a wide range of philosophical debates. The concept is used in debates around, e.g., explanation, scientific realism and the philosophy of mind, to name just a few. Many of those who engage in these debates share some kind of physicalism, that is, the view that the fundamental reality is only physical, and therefore prima facie should care whether there is causation in physics. Consequently, a contrary result to the one reached here would have strongly undermined the use of the term 'causation' within all these debates. If, on the other hand, the conclusion of the present work is correct, then this can be seen as an invitation to carry over the result into other fields, for example, by exploring how $\mathrm{CQT}_{\mathrm{QFT}}$ can fit into a coherent scientific realism.

Furthermore, $\mathrm{CQT}_{\mathrm{QFT}}$ enhances our understanding of QFT. It arguably makes a difference to know that processes in QFT are causal processes, as opposed to mere correlations. On the one hand, causal relations have a practical impact that is emphasised by the possibility to manipulate the effect by manipulating the cause. On the other hand, given that we experience our everyday world as causal, a causal interpretation of QFT brings the latter closer to our everyday world. On the assumption that we understand our everyday world better than QFT, we therefore gain a more intelligible picture of the latter. This is true even if the forms of causation in physics and everyday life are different in some respects, as at least they will share most intuitions, or else they could not both be called causation.

I see this advance in the understanding of QFT as a major advantage of the causal interpretation over other interpretations of QFT. There is now an extensive and inconclusive debate on what QFT tells us about the world. Large parts of this debate are concerned with the right ontology for QFT. To the contrary, the causal interpretation of QFT does not need any controversial ontological interpretation. It rests solely on general and well founded theoretical features of QFT. The causal interpretation therefore advances our understanding of QFT, but avoids some yet unresolved problems. At the same time, of course, the causal interpretation cannot be regarded as a complete interpretation of QFT.

This brings me to the question of how the account of causation presented here could be developed further. An interesting project would be to embed $\mathrm{CQT_{QFT}}$ into a broader metaphysics of properties. A very general question in this respect is the compatibility with Humean or dispositional properties. Furthermore, it would be interesting to investigate where $\mathrm{CQT_{QFT}}$ sits with respect to broader ontological positions like ontic structural realism. This is especially the case as it is sometimes claimed that ontic structural realism is incompatible with causation. Having a concrete theory of physical causation like $\mathrm{CQT_{QFT}}$ could help to push this debate forward.

On the other hand, I am aware that I have not covered all physical phenomena that might have an impact on causation. I have already mentioned the Aharonov-Bohm and the Casimir effects earlier. Also, my discussion of QFT was mostly confined to quantum electrodynamics and here an extension to other forces would be desirable. Furthermore, there are still open questions in physics that are important to the present project. I have discussed the measurement problem and entanglement at length, but there is more. The biggest issue is certainly that we have good reasons to assume that QFT is not the final quantum theory, given its incompatibility with the general theory of relativity. At the moment, one can only speculate how a grand unifying theory would look like, but it might change our view of the world considerably. The only comment I can make to this is that, quite trivially, all what I have said about physics is conditional on further developments in physics.

I wish to end this work with a final comment on a potential worry. I have argued that our intuitions on causation are realised by certain traits of QFT. As it turned out, our intuitions and QFT is a really good fit; maybe too good. Intuitions about a concept are always unsharp and there can be a suspicion that I tailored them in a way to comply with QFT. This, in effect, would have made the whole work a self fulfilling prophecy. However, is it really surprising that our intuitions about causation work so well with QFT, but not so much with classical mechanics or quantum mechanics? There are good reasons for assuming that QFT, even though not the final theory, is more correct than QM and even more than classical mechanics. Not least because it is covariant, and we have very high confidence in STR. Consequently, by locality, spacelike separated events cannot influence each other and conserved quantities cannot simply pop up out of nowhere. This is in agreement with our everyday life, which arguably has formed our intuitions on causation to a large extent. If instead, we lived in a world that was not relativistic, then presumably we would observe completely different processes, and had a different concept of causation. For example, using nuclear fission to produce energy would presumably not be possible in a world in which STR was false. Now, if the fundamental physical world determines our everyday experiences, and QFT is at least approximately true on energy levels that are most relevant for this determination, then it cannot be surprising that a very similar concept of causation works on both levels. Our intuitions about causation agree so well with QFT, because we live in a world that is correctly described by QFT.[1]

[1]This work is not aimed at everyday causation and I will therefore not develop this theme any further. However, Elga (2007) explores the thought in more detail.

Bibliography

Ahmed, A. (2007). Agency and causation. In H. Price & R. Corry (Eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford: Oxford University Press.

Aiello, M., Castagnino, M., & Lombardi, O. (2007). The arrow of time: from universe time-asymmetry to local irreversible processes. arXiv: gr-qc/0608099v2.

Albert, D. (1994). *Quantum Mechanics and Experience*. Cambridge (Massachusetts): Harvard University Press, 2. edition.

Albert, D. (2000). *Time and Chance*. Cambridge (Massachusetts): Harvard University Press.

Aristotle (1930). *Physics*. Oxford: Oxford University Press. Translated by R. P. Hardie and R. K. Gaye.

Armstrong, D. M. (2004). Going through the open door again: Counterfactual versus singularist theories of causation. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and Counterfactuals*. Cambridge (Massachusetts): The MIT Press.

Aronson, J. L. (1971). On the grammar of 'cause'. *Synthese*, 22, 414–430.

Atkinson, D. (2006). Does quantum electrodynamics have an arrow of time? *Studies in History and Philosophy of Modern Physics*, 37, 528–541.

Auyang, S. Y. (1995). *How is Quantum Field Theory Possible?* Oxford: Oxford University Press.

Bacciagaluppi, G. (2012). The role of decoherence in quantum mechanics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Winter 2012 edition.

Banks, T., Susskind, L., & Peskin, M. E. (1984). Difficulties for the evolution of pure states into mixed states. *Nuclear Physics B*, 244, 125–134.

Barrett, J. A. (2000). On the nature of measurement records in relativistic Quantum Field Theory. http://philsci-archive.pitt.edu/197/.

Bartels, A. & Wohlfarth, D. (2014). How fundamental physics represents causality. In M. C. Galavotti, D. Dieks, W. J. Gonzalez, S. Hartmann, T. Uebel, & M. Weber (Eds.), *New Directions in the Philosophy of Science*, volume 5 of *The Philosophy of Science in a European Perspective* (pp. 485–500). Springer International Publishing.

Beebee, H. (2004a). Causing and nothingness. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Counterfactuals and Causation* (pp. 291–308). Cambridge (Massachusetts): The Mit Press.

Beebee, H. (2004b). Chance-changing causal processes. In P. Dowe & P. Noordhof (Eds.), *Cause and Chance. Causation in an Indeterministic World*. London: Routledge.

Beebee, H. (2007). Hume on causation: The projectivist interpretation. In H. Price & R. Corry (Eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford: Oxford University Press.

Bell, J. S. (1987). *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press.

Bell, J. S. (1990). Against 'measurement'. *Physics World*, (pp. 33–40).

Berkovitz, J. (2014). Action at a distance in quantum mechanics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Spring 2014 edition.

Bertlmann, R. A., Durstberger, K., & Hiesmayr, B. C. (2003). Decoherence of entangled kaons and its connection to entanglement measures. *Physical Review A*, 68.

Bertlmann, R. A., Grimus, W., & Hiesmayr, B. C. (2002). The EPR paradox in massive systems or about strange particles. In R. A. Bertlmann & A. Zeilinger (Eds.), *Quantum [Un]speakables: From Bell to quantum information* chapter 11. Berlin: Springer Verlag.

Blanchard, P., Jakóbczyk, L., & Olkiewicz, R. (2001). Measures of entanglement based on decoherence. *Journal of Physics A: Mathematical and General*, 34, 8501–8516.

Blondeau, J. & Ghins, M. (2012). Is there an intrinsic criterion for causal lawlike statements? *International Studies in the Philosophy of Science*, 26(4), 381–401.

Bontly, T. D. (2006). What is an empirical analysis of causation? *Synthese*, 151(2), 177–200.

Braddon-Mitchell, D. & Nola, R. (2009). Introducing the canberra plan. In D. Braddon-Mitchell & R. Nola (Eds.), *Conceptual Analysis and Philosophical Naturalism*. Cambridge (Massachusetts): The MIT Press.

Brading, K. & Brown, H. R. (2003). Symmetries and Noether's theorems. In K. Brading & E. Castellani (Eds.), *Symmetries in Physics. Philosophical Reflections*. Cambridge: Cambridge University Press.

Bransden, B. H. & Joachain, C. J. (2000). *Quantum Mechanics*. Harlow (England): Pearson Education, 2. edition.

Breuer, H.-P. & Petruccione, F. (2002). *The Theory of Open Quantum Systems*. Oxford: Oxford University Press.

Brown, H. R. (1986). The insolubility proof of the quantum measurement problem. *Foundations of Physics*, 16(9), 857–870.

Brown, H. R. (2007). *Physical Relativity: Space-time Structure from a Dynamical Perspective*. Oxford: Oxford University Press.

Brune, M., Hagley, E., Dreyer, J., Maître, X., Maali, A., Wunderlich, C., Raimond, J. M., & Haroche, S. (1996). Observing the progressive decoherence of the "meter" in a quantum measurement. *Physical Review Letters*, 77(24).

Bub, J. & Pitowsky, I. (2010). Two dogmas about quantum mechanics. In S. Saunders, J. Barrett, A. Kent, & D. Wallace (Eds.), *Many Worlds. Everett, Quantum Theory and Reality*. Oxford: Oxford University Press.

Bueno, O. & French, S. (2011). How theories represent. *British Journal for the Phillosophy of Science*, 62, 857–894.

Busch, P. (2003). The role of entanglement in quantum measurement and information processing. *International Journal of Theoretical Physics*, 42, 937–941. arXiv: quant-ph/0209090v3.

Busch, P., Lathi, P. J., & Mittelstaed, P. (1996). *Quantum Theory of Measurement*. Berlin, Heidelberg: Springer, 2. edition.

Butterfield, J. (1992a). Bell's theorem: What it takes. *British Journal for the Philosophy of Science*, 43(1), 41–83.

Butterfield, J. (1992b). David Lewis meets John Bell. *Philosophy of Science*, 59(1), 26–43.

Butterfield, J. (1998). Determinism and indeterminism. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*, volume 3 (pp. 33–39). New York: Routledge.

Butterfield, J. (2007). Reconsidering relativistic causality. *International Studies in the Philosophy of Science*, 21(3), 295–328.

Callender, C. (1997). What is 'the problem of the direction of time'? *Philosophy of Science*, 64, 223–234. Supplement. Proceedings of the 1996 Biennal Meetings of the Philosophy of Science Association. Part II: Symposia Papers.

Callender, C. (1998). The view from no-when. *British Journal for the Phillosophy of Science*, 49, 135–159.

Callender, C. (2004). There is no puzzle about the low-entropy past. In C. Hitchcock (Ed.), *Contemporary debates in philosophy of science*. Malden: Blackwell Publishing.

Cao, T. Y. (1997). *Conceptual developments of 20th century field theories*. Cambridge: Cambridge University Press.

Caro, M. D. & Macarthur, D. (2004). Introduction: The nature of Naturalism. In M. D. Caro & D. Macarthur (Eds.), *Naturalism in Question*. Cambridge, Massachusetts: Harvard University Press.

Carroll, J. W. (2009). Ant-reductionism. In H. Beebee, P. Menzies, & C. Hitchcock (Eds.), *The Oxford Handbook of Causation*. Oxford: Oxford University Press.

Cartwright, N. (1978). The only real probabilities in quantum mechanics. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, One: Contributed Papers, 54–59.

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

Cartwright, N. (1980). Measuring position probabilities. In *Studies in the foundations of quantum mechanics*. East Lansing (Michigan): Philosophy of Science Association.

Castagnino, M. & Lombardi, O. (2009). The global non-entropic arrow of time: from global geometrical asymmetry to local energy flow. *Synthese*, 169, 1–25.

Chakravartty, A. (2013). On the prospects of naturalized metaphysics. In D. Ross, J. Ladyman, & H. Kincaid (Eds.), *Scientific Metaphysics*. Oxford: Oxford University Press.

Chalmers, D. (2013). Intuitions in philosophy: A minimal defense. http://consc.net/papers/intuition.pdf.

Chisholm, R. M. (1992). The basic ontological categories. In K. Mulligan (Ed.), *Language, Truth and Ontology* (pp. 1–13). Dordrecht: Kluwer.

Chudnoff, E. (2011). The nature of intuitive justification. *Philosophical Studies*, 153(2), 313–333.

Collingwood, G. (1940). *An Essay in Metaphysics*. Oxford: Oxford University Press.

Collins, J., Hall, N., & Paul, L. A. (2004). Counterfactuals and causation: History, problems, and prospects. In *Causation and Counterfactuals*. Cambridge, Massachusetts: The MIT Press.

Corry, R. (2015). Retrocausal models for EPR. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 49, 1–9.

da Costa, N. & French, S. (2003). *Science and Partial Truth*. New York: Oxford University Press.

de Riedmatten, H., Laurat, J., Chou, C., Schomburg, E., Felinto, D., & Kimble, H. (2006). Direct measurement of decoherence for entanglement between a photon and stored atomic excitation. *Physical Review Letters*, 97(11).

D'Espagnat, B. (1966). Two remarks on the theory of measurement. *Il Nuovo Cimento, Supplement 4*, (pp. 828–38).

Devitt, M. (1984). *Realism and Truth.* Princeton: Princeton University Press.

Devitt, M. (2005). There is no *a Priori.* In M. Steup & E. Sosa (Eds.), *Contemporary Debates in Epistemology.* Malden: Blackwell Publishers.

Dieks, D. (2001). Space and time in particle and field physics. *Studies in History and Philosophy of Modern Physics*, 32(2), 217–241.

Dowe, P. (1992). Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of Science*, 59(2), 195–216.

Dowe, P. (1995). Causality and conserved quantities: A reply to Salmon. *Philosophy of Science*, 62(2), 321–333.

Dowe, P. (2000). *Physical Causation.* Cambridge: Cambridge University Press, 1. edition.

Dowe, P. (2004). Chance-lowering causes. In P. Dowe & P. Noordhof (Eds.), *Cause and Chance. Causation in an Indeterministic World.* London: Routledge.

Ducasse, C. (1926). On the nature and the observability of the causal relation. *Journal of Philosophy*, 23, 57–68.

Duetsch, M. & Fredenhagen, K. (2000). Perturbative algebraic field theory, and deformation quantization. *Fields Inst. Commun.*, (pp. 151–160).

Duncan, A. (2012). *The Conceptual Framework of Quantum Field Theory.* Oxford: Oxford University Press.

Durt, T. (2004). Quantum entanglement, interaction, and the classical limit. arXiv: quant-ph/0401121v1.

Eagle, A. (2007). Pragmatic causation. In H. Price & R. Corry (Eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited.* Oxford University Press.

Earman, J. (1974). An attempt to add a little direction to "The problem of the Direction of Time". *Philosophy of Science*, 41(1), 15–47.

Earman, J. (1976). Causation: A matter of life and death. *The Journal of Philosophy*, 73(1), 5–25.

Earman, J. (1986). *A primer on determinism*. Dordrecht: D. Reidel.

Earman, J. (2002). What time reversal invariance is and why it matters. *International Studies in the Philosophy of Science*, 16(3), 245–264.

Earman, J. (2007). Aspects of determinism in modern physics. In J. Butterfield & J. Earman (Eds.), *Philosophy of Physics, Part B* (pp. 1369–1434). Amsterdam: North-Holland.

Earman, J. (2011). Sharpening the electromagnetic arrow(s) of time. In C. Callender (Ed.), *The Oxford handbook on time*. Oxford: Oxford University Press.

Earman, J. & Fraser, D. (2006). Haag's theorem and its implications for the foundations of quantum field theory. *Erkenntnis*, 64(3), 305–344.

Earman, J. & Shimony, A. (1968). A note on measurement. *Il Nuovo Cimento*, 54(2), 332–334.

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

Einstein, A. (1948). Quanten-Mechanik und Wirklichkeit. *Dialectica*, 2, 320Ð324.

Einstein, A., Podolsky, B., & Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, (47), 777–780.

Elby, A. (1992). Should we explain the EPR correlations causally? *Philosophy of Science*, 59(1), 16–25.

Elga, A. (2007). Isolation and folk physics. In Huw Price & Richard Corry (Eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford: Oxford University Press.

Emary, C. & Beenakker, C. W. J. (2004). Relation between entanglement measures and bell inequalities for three qubits. *Physical Review A*, 69. 032317.

Esfeld, M. (2001). Lewis' causation and quantum correlations. In W. Spohn, M. Ledwig, & M. Esfeld (Eds.), *Current Issues in Causation*. Parderborn: Mentis.

Esfeld, M. (2004). Quantum entanglement and a metaphysics of relations. *Studies in History and Philosophy of Modern Physics*, 35, 601–617.

Facchi, P., Tasaki, S., Pascazio, S., Nakazato, H., Tokuse, A., & Lidar, D. A. (2005). Control of decoherence: Analysis and comparison of three different strategies. *Physical Review A*, 71.

Fair, D. (1979). Causation and the flow of energy. *Erkenntnis*, 14, 219–250.

Falkenburg, B. (2007). *Particle Metaphysics: A critical account of subatomic reality*. Berlin: Springer-Verlag.

Farr, M. & Reutlinger, A. (2013). A relic of a bygone age? Causation, time symmetry and the directionality argument. *Erkenntnis*, 78(2), 215–235.

Fenton-Glynn, L. & Kroedel, T. (2015). Relativity, quantum entanglement, counterfactuals, and causation. *British Journal for the Philosophy of Science*, 66(1), 45–67.

Feynman, R. (1985). *QED: The strange theory of light and matter.* Princeton: Princeton University Press.

Field, H. (2003). Causation in a physical world. In M. Loux & D. Zimmerman (Eds.), *Oxford handbook of metaphysics.* Oxford: Oxford University Press.

Fine, A. (1969). On the general quantum theory of measurement. In *Proceedings of the Cambridge Philosophical Society*, volume 65 (pp. 111–122).

Fine, A. (1970). Insolubility of the quantum measurement problem. *Physical Review D*, 2(12), 2783–2787.

Fleming, G. (2002). Comments on Paul Teller's book, "An Interpretive Introduction to Quantum Field Theory". In M. Kuhlmann, H. Lyre, & A. Wayne (Eds.), *Ontological Aspects of Quantum Field Theory* (pp. 135–144). New Jersey: World Scientific.

Fleming, G. & Butterfield, J. N. (1999). Strange positions. In J. N. Butterfield & C. Pagonis (Eds.), *From Physics to Philosophy.* Cambridge: Cambridge University Press.

Fox, T. (2008). Haunted by the spectre of virtual particles: A philosophical reconsideration. *Journal for General Philosophy of Science*, 39(1), 35–51.

Fraser, D. (2008). The fate of 'particles' in quantum field theories with interactions. *Studies in History and Philosophy of Modern Physics*, 39, 841–859.

Fraser, D. (2009). Quantum field theory: Underdetermination, inconsistency, and idealization. *Philosophy of Science*, 76(4), 536–567.

Fraser, D. (2011). How to take particle physics seriously: A further defence of axiomatic quantum field theory. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 42(2), 126–135.

Fraser, D. L. (2006). *Haag's theorem and the interpretation of quantum field theories with interactions.* PhD thesis, University of Pittsburgh.

French, S. (1989). Individuality, supervenience and Bell's theorem. *Philosophical Studies*, 55(1), 1–22.

French, S. (1999). Models and mathematics in physics: the role of group theory. In J. Butterfield & C. Pagonis (Eds.), *From physics to philosophy.* Cambridge: Cambridge University Press.

French, S. (2014). *The Structure of the World*. Oxford: Oxford University Press.

French, S. & Krause, D. (2006). *Identity in physics: a historical, philosophical, and formal analysis*. Oxford: Clarendon Press.

French, S. & Ladyman, J. (2003). Remodelling structural realism: Quantum physics and the metaphysics of structure. *Synthese*, 136(1), 31–56.

Frisch, M. (2000). (Dis-)solving the puzzle of the arrow of radiation. *British Journal for the Philosophy of Science*, 51, 381–410.

Frisch, M. (2005). Counterfactuals and the past hypothesis. *Philosophy of Science*, 72(5), 739–750.

Frisch, M. (2009a). Causality and dispersion: A reply to John Norton. *British Journal for the Phillosophy of Science*, 60, 487–495.

Frisch, M. (2009b). 'The most sacred tenet'? Causal reasoning in physics. *British Journal for the Phillosophy of Science*, 60, 459–474.

Frisch, M. (2010). Causal models and the asymmetry of state preparation. In M. Suárez, M. Dorato, & M. Rédei (Eds.), *EPSA Philosophical Issues in the Sciences* (pp. 75–85). Springer Netherlands.

Frisch, M. (2012). No place for causes? Causal skepticism in physics. *European Journal for Philosophy of Science*, 2(3), 313–336.

Frisch, M. (2014). *Causal Reasoning in Physics*. Cambridge: Cambridge University Press.

Garbaczewski, P. & Olkiewicz, R., Eds. (2002). *Dynamics of dissipation*. Berlin: Springer.

Gasking, D. (1955). Causation and recipes. *Mind*, 64, 479–478.

Gemmer, J. & Mahler, G. (2001). Entanglement and the factorization-approximation. *European Physical Journal D*, 17, 385–393. arXiv: quant-ph/0109140v1.

Georgi, H. (1999). *Lie Algebras in Particle Physics*. Westview Press, 2. edition.

Gillies, D. (2000). *Philosophical Theories of Probability*. London: Routledge.

Gillies, D. (2001). Critical notices-judea pearl, causality: Models, reasoning, and inference. *British Journal for the Philosophy of Science*, 52(3), 613–622.

Gilmore, R. (1974). *Lie Groups, Lie Algebras, and some of their Applications*. New York: John Wiley & Sons.

Giulini, D., Joos, E., Kiefer, C., Kupsch, J., Stamatescu, I.-O., & Zeh, H. D. (1996). *Decoherence and the Appearance of a Classical World in Quantum Theory*. Berlin, Heidelberg: Springer.

Godfrey-Smith, P. (2009). Causal pluralism. In H. Beebee, P. Menzies, & C. Hitchcock (Eds.), *The Oxford Handbook of Causation*. Oxford: Oxford University Press.

Gold, T. (1966). Cosmic processes and the nature of time. In R. Colodny (Ed.), *Mind and Cosmos*. Pittsburgh: University of Pittsburgh Press.

Good, I. J. (1961). A causal calsulus I-II. *British Journal for the Phillosophy of Science*, 11 & 12, 305–318 & 43–51.

Goswami, S. & Ota, T. (2008). Testing non-unitarity of neutrino mixing matrices at neutrino factories. arXiv: 0802.1434v1 [hep-ph].

Gracia, J. J. E. (1988). *Individuality: An Essay on the Foundations of Metaphysics*. Albany: State University of New York Press.

Greaves, H. & Wallace, D. (2011). Empirical consequences of symmetries. ArXiv: 1111.4309 [physics.hist-ph].

Green, D. (2000). *The physics of particle detectors*. Cambridge: Cambridge University Press.

Greiner, W. & Reinhardt, J. (1996). *Field Quantization*. Berlin: Springer-Verlag.

Greiner, W. & Reinhardt, J. (2009). *Quantum Electrodynamics*. Berlin: Springer-Verlag.

Grupen, C. & Schwartz, B. (2008). *Particle Detectors*, volume 26 of *Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology*. Cambridge: Cambridge University Press, 2 edition. with contributions from H. Spieler, S. Eidelman and T. Stroh.

Haag, R. (1996). *Local Quantum Physics. Fields, Particles, Algebras*. Berlin: Springer, 2. edition.

Haag, R. & Swieca, J. A. (1965). When does a quantum field theory describe particles? *Communications in Mathematical Physics*, 1(4), 308–320.

Hall, N. (2004a). Causation and the price of transitivity. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and Counterfactuals*. Cambridge (Massachusetts): The MIT Press.

Hall, N. (2004b). Two concepts of causation. In J. Collins, N. Hall, & L. Paul (Eds.), *Causation and Counterfactuals*. Cambridge (Massachusetts): The Mit Press.

Hall, N. (2007). Structural equations and causation. *Philosophical Studies*, 132(1), 109–136.

Hall, N. (2011). Causation and the sciences. In S. French & J. Saatsi (Eds.), *The Continuum Companion to the Sciences*. London: Continuum Press.

Halvorson, H. (2001). Reeh-Schlieder defeats Newton-Wigner: on alternative localisation schemes in relativistic quantum field theory. *Philosophy of Science*, 68, 111–133.

Halvorson, H. & Clifton, R. (2002). No place for particles in relativistic quantum theories? *Philosophy of Science*, 69(1), 1–28.

Harré, R. (1988). Parsing the amplitudes. In H. R. Brown & R. Harré (Eds.), *Philosophical Foundations of Quantum Field Theory*. Oxford: Clarendon Press.

Hartle, J. B. (1994). Unitarity and causality in generalized quantum mechanics for nonchronal space-times. *Physical Review*, D49, 6543–6555. arXiv: gr-qc/9309012.

Hausman, D. M. (1998). *Causal Asymmetries*. Cambridge: Cambridge University Press.

Hawking, S. W. (1982). The unpredictability of quantum gravity. *Communications in Mathematical Physics*, 87, 395–415.

Healey, R. (2009). Holism and nonseparability in physics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Spring 2009 edition.

Healey, R. A. (2014). Causality and chance in relativistic quantum field theories. *Studies in History and Philosophy of Modern Physics*. http://dx.doi.org/10.1016/j.shpsb.2014.03.002i.

Heathcote, A. (1989). A theory of causality: Causality = interaction (as defined by a suitable quantum field theory). *Erkenntnis*, 31, 77–108.

Heidelberger, M. (1992). Kausalität: Eine Problemübersicht. *Neue Hefte für Philosophie*, 32/33, 130–153.

Heidelberger, M. (2010). Functional relations and causality in Fechner and Mach. *Philosophical Psychology*, 23(2), 163–172.

Hill, C. E. (1994). Ion and electron sources. http://cdsweb.cern.ch/record/276893/files/ps-94-036.pdf.

Hintikka, J. (1999). The emperor's new intutions. *Journal of Philosophy*, 96(3), 127–147.

Hitchcock, C. (2003). Of humean bondage. *British Journal for the Phillosophy of Science*, 54(1), 1–25.

Hitchcock, C. (2004a). Do all and only causes raise the probabilities of effects? In J. Collind, N. Hall, & L. A. Paul (Eds.), *Causation and Counterfactuals*. Cambridge (Massachusetts): MIT Press.

Hitchcock, C. (2004b). Routes, processes and chance-lowering causes. In P. Dowe & P. Noordhof (Eds.), *Cause and Chance. Causation in an Indeterministic World*. London: Routledge.

Hitchcock, C. (2009). Structural equations and causation: six counterexamples. *Philosophical Studies*, 144(3), 391–401.

Hitchcock, C. R. (1995). Salmon on explanatory relevance. *Philosophy of Science*, 62(2), 304–320.

Hitchcock, C. R. (1996). The role of contrast in causal and explanatory claims. *Synthese*, 107(3), 395–419.

Hitchcock, C. R. (2007). What Russell got right. In H. Price & R. Corry (Eds.), *Causation, Physics, and the Constitution of Reality. Russell's Republic revisited* (pp. 45–65). Oxford: Clarendon Press.

Hoefer, C. (2004). Causality and determinism: Tension, or outright conflict? *Revista de Filosophía*, 29(2), 99–115.

Hoefer, C. (2010). Causal determinism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Spring 2010 edition.

Hofer-Szabó, G., Rédei, M., & Szabó, L. E. (2013). *The Principle of the Common Cause*. Cambridge: Cambridge University Press.

Holman, M. (2008). On arguments for linear quantum dynamics. arXiv: quant-ph/0612209v1.

Hornberger, K. (2009). Introduction to decoherence theory. In C. V. A. Buchleitner & M. Tiersch (Eds.), *Entanglement and Decoherence. Foundations and Modern Trends*, volume 768 of *Lecture Notes in Physics*. Berlin: Springer.

Hornberger, K., Uttenthaler, S., Brezger, B., Hackermüller, L., Arndt, M., & Zeilinger, A. (2003). Collisional decoherence observed in matter wave interferometry. *Physical Review Letters*, 90(16).

Horodecki, R., Horodecki, P., Horodecki, M., & Horodecki, K. (2009). Quantum entanglement. *Reviews of Modern Physics*, 81, 865–942.

Howard, D. (1989). Holism, separability and the metaphysical implications of the bell experiments. In J. Cushing & E. McMullin (Eds.), *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem* (pp. 224–253). Notre Dame (Indiana): University of Notre Dame Press.

Huggett, N. (2000). Philosophical foundations of quantum field theory. *The British Journal for the Philosophy of Science*, 51, 617–637.

Hunt, I. (2005). Omission and prevention as cases of genuine causation. *Philosophical Papers*, 34(2), 209–233.

Jackson, F. (1998). *From Metaphysics to Ethics. A Defense of Coneptual Analysis*. Oxford: Clarendon Press.

Jaffe, A. (2007). Quantum theory and relativity.

Jaffe, R. L. (2005). The casimir effect and the quantum vacuum. *Phys. Rev. D*, 72. arXiv: hep-th/0503158.

Kaiser, D. (2005). *Drawing Theories Apart. The Dispersion of Feynman Diagrams in Postwar Physics*. Chicago: The University of Chicago Press.

Kennard, E. H. (1931). Quantum-mechanical motion of free electrons in electromagnetic fields. *Proceedings of the National Academy of Sciences of the United States of America*, 17(1), 58 – 62.

Kiefer, C. (1992). Decoherence in quantum electrodynamics and quantum gravity. *Physical Review D*, 46(4), 1658–1670.

Kiess, T. E., Shih, Y. H., Sergienko, A. V., & Alley, C. O. (1995). Tunable bell-inequality violations by non-maximally-violating states in type-ii parametric down-conversion. *Physical Review A*, 52(4), 3344–3347.

Kingsbury, J. & McKeown-Green, J. (2009). Jackson's armchair: The only chair in town? In D. Braddon-Mitchell & R. Nola (Eds.), *Conceptual Analysis and Philosophical Naturalism*. Cambridge, Massachusetts: The MIT Press.

Kistler, M. (2006). *Causation and Laws of Nature*. London: Routledge.

Kitcher, P. (1992). The naturalists return. *Philosophical Review*, 101(1), 53–114.

Knight, J. M. (1961). Strict localization in quantum field theory. *Journal of Mathematical Physics*, 2(4), 459–471.

Kolmogorov, A. N. (1956). *Foundations of the Theory of Probability*. New York: Chelsea Phublishing Company, 2. edition.

Krips, H. (1989). The objectivity of quantum probabilities. *Australasian Journal of Philosophy*, 67(4), 423–431.

Kuhlmann, M., Lyre, H., & Wayne, A., Eds. (2002). *Ontological Aspects of Quantum Field Theory*. New Jersey: World Scientific.

Kupsch, J. (2000). Mathematical aspects of decoherence. In P. Blanchard, D. Giulini, E. Joos, C. Kiefer, & J.-O. Stamatescu (Eds.), *Decoherence; Theretical, Experimental, and Conceptual Problems. Proceedings of a Workshop Held in Bielefeld, Germany, 10-14 November 1998*. Berlin, Heidelberg: Springer.

Kutach, D. (2007). The physical foundations of causation. In H. Price & R. Corry (Eds.), *Causation, Physics, and the Constitution of Reality*. Oxford: Oxford University Press.

Kutach, D. (2010). Empirical analysis of causation. http://sagaciousmatter.org /KutachEmpiricalAnalysesOfCausation.pdf.

Kutach, D. (2011). The asymmetry of influence. In C. Callender (Ed.), *Oxford Handbook of Philosophy of Time*. Oxford: Oxford University Press.

Ladyman, J. & Ross, D. (2007). *Every Thing Must Go. Metaphysics Naturalized*. Oxford: Oxford University Press.

Landsman, N. & Reuvers, R. (2013). A flea on Schrödinger's cat. *Foundations of Physics*, 43(3), 373–407.

Larsson, J.-A., Giustina, M., Kofler, J., Wittmann, B., Ursin, R., & Ramelow, S. (2014). Bell-inequality violation with entangled photons, free of the coincidence-time loophole. *Physical Review A*, 90.

Lewis, D. (1970). How to define theoretical terms. *Journal of Philosophy*, 67(13), 427–446.

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50(3), 249–258.

Lewis, D. (1973). Causation. *Journal of Philosophy*, 70, 556–567.

Lewis, D. (1986a). Causal explanation. In *Philosophical Papers*, volume II. Oxford: Oxford University Press.

Lewis, D. (1986b). Causation. In *Philosophical Papers*, volume II. New York, Oxford: Oxford University Press.

Lewis, D. (1986c). Events. In *Philosophical Papers*, volume II. Oxford: Oxford University Press.

Lewis, D. (2004a). Causation as influence. In J. Collind, N. Hall, & L. A. Paul (Eds.), *Causation and Counterfactuals*. Cambridge (Massachusetts): MIT Press.

Lewis, D. (2004b). Void and object. In J. Collind, N. Hall, & L. A. Paul (Eds.), *Causation and Counterfactuals*. Cambridge (Massachusetts): MIT Press.

LHC-Experiments-Committee (1999). Atlas detector and physics performance: Technical design report, 1. http://cdsweb.cern.ch/record/391176.

Li, M., Fei, S.-M., & Li-Jost, X. (2010). Quantum entanglement: Separability, measure, fidelity of teleportation, and distillation. *Advances in Mathematical Physics*. Article ID 301072.

Liebesman, D. (2011). Causation and the Canberra plan. *Pacific Philosophical Quarterly*, 92, 232–242.

Lowe, E. J. (2006). *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*. Oxford: Oxford University Press.

Lowe, E. J. (2010). Ontological dependence. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Spring 2010 edition.

Lupher, T. (2009). A physical critique of physical causation. *Synthese*, 167(1), 67–80.

Lyre, H. (2010). Why quantum theory is possibly wrong. *Foundations of Physics*, 40(9), 1429–1438.

Lyre, H. (2012). Structural invariants, structural kinds, structural laws. In D. Dieks, W. J. Gonzalez, S. Hartmann, M. Stöltzner, & M. Weber (Eds.), *Probabilities, Laws, and Structures*. Dordrecht: Springer.

Mach, E. (1886). *Analyse der Empfindungen*. Jena: Fischer, 5th edition.

Maciejko, J. (unpublished). Representations of Lorentz and Poincaré groups. http://einrichtungen.physik.tu-muenchen.de/T30f/lec/QFT/groups.pdf.

Mackie, J. L. (1980). *The Cement of the Universe: A Study of Causation*. Oxford: Clarendon Press.

Maggiore, M. (2005). *A Modern Introduction to Quantum Field Theory*. Oxford: Oxford University Press.

Malament, D. B. (1996). In defense of dogma: Why there cannot be a relativistic quantum mechanics of (localizable) particles. In R. Clifton (Ed.), *Perspectives on quantum reality : Non-relativistic, relativistic, and field-theoretic* (pp. 1–10). Dordrecht: Kluwer Academic Publishers.

Maudlin, T. (1995). Three measurement problems. *Topoi*, 14, 7–15.

Maudlin, T. (2002). *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics*. Malden (Massachusetts): Blackwell Publishers, second edition edition.

Maxwell, N. (1976). Toward a micro realistic version of quantum mechanics. part i, ii. *Foundations of Physics*, 6(3, 6), 275–292, 661–676.

Maxwell, N. (1982). Instead of particles and fields: A micro realistic quantum "smearon" theory. *Foundations of Physics*, 12(6), 607–631.

Maxwell, N. (1988). Quantum propensiton theory: A testable resolution of the wave/particle dilemma. *British Journal for the Phillosophy of Science*, 39, 1–50.

McGrath, S. (2005). Causation by omission: A dilemma. *Philosophical Studies*, 123, 125–148.

Mellor, D. H. (1987). The singularly affecting facts of causation. In P. P. et al. (Ed.), *Metaphysics and Morality: Essays in Honour of J.J.C. Smart*. Oxford: B. Blackwell.

Mellor, D. H. (1995). *The Facts of Causation*. London: Routledge.

Menzies, P. (1988). Against causal reductionism. *Mind*, 97(388), 551–574.

Menzies, P. (1989). Probabilistic causation and causal processes: A critique of Lewis. *Philosophy of Science*, 56(4), 642–663.

Menzies, P. (1996). Probabilistic causation and the pre-emption problem. *Mind*, 105(417), 85–117.

Menzies, P. (1999). Intrinsic versus extrinsic conceptions of causation. In H. Sankey (Ed.), *Causation and laws of nature* (pp. 313–329). Kluwer Academic Publishers.

Menzies, P. (2004). Difference-making in context. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Counterfactuals and Causation* (pp. 139–180). Cambridge (Massachusetts): The Mit Press.

Menzies, P. (2009). Platitudes and counterexamples. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *The Oxford Handbook of Causation* (pp. 341–367). Oxford: Oxford University Press.

Menzies, P. (2014). Counterfactual theories of causation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Spring 2014 edition.

Menzies, P. & Price, H. (1993). Causation as a secondary quality. *British Journal for the Phillosophy of Science*, 44, 187–203.

Meynell, L. (2008). Why feynman diagrams represent. *International Studies in the Philosophy of Science*, 22(1), 39–59.

Mirman, R. (1975). The direction of time. *Foundations of Physics*, 5(3), 491–511.

Munro, W. J., Nemoto, K., & White, A. G. (2001). The Bell inequality: A measure of entanglement? *journal of Modern Optics*, 48(7), 1239–1246.

Nagashima, Y. (2010). *Elementary Particle Physics. Volume 1: Quantum Field Theory and Particles*. Weinheim: WILEY-VCH Verlag.

Näger, P. M. (2013). Causal graphs for EPR experiments. http://philsci-archive.pitt.edu/9915/.

Ney, A. (2009). Physical causation and difference-making. *British Journal for the Philosophy of Science*, 60(4), 737–764.

Nolan, D. (2009). Platitudes and metaphysics. In D. Braddon-Mitchell & R. Nola (Eds.), *Conceptual Analysis and Philosophical Naturalism*. Cambridge (Massachusetts): The MIT Press.

Norton, J. (2008). The dome: An unexpectedly simple failure of determinism. *Philosophy of Science*, 75(5), 786–798.

Norton, J. D. (2007). Causation as folk science. In H. Price & R. Corry (Eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford University Press.

Norton, J. D. (2009). Is there an independent principle of causality in physics? *British Journal for the Phillosophy of Science*, 60, 475–486.

Papineau, D. (2009). Naturalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Spring 2009 edition.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Peskin, M. E. & Schroeder, D. V. (1995). *An Introduction to Quantum Field Theory*. Reading (Massachusetts): Perseus Books.

Plenio, M. B. & Virmani, S. (2007). An introduction to entanglement measures. *Quantum Information & Computation*, 7, 1–51.

Price, H. (1991). Agency and probabilistic causality. *British Journal for the Phillosophy of Science*, 42, 157–176.

Price, H. (1996a). Backward causation and the direction of causal processes: Reply to Dowe. *Mind*, 105(419), 467–474.

Price, H. (1996b). *Time's arrow and Archimedes' point*. New York, Oxford: Oxford University Press.

Price, H. (2004). On the origins of the arrow of time: Why there is still a puzzle about the low-entropy past. In C. Hitchcock (Ed.), *Contemporary debates in philosophy of science*. Malden: Blackwell Publishing.

Price, H. (2011). The flow of time. In C. Callender (Ed.), *The Oxford handbook of philosophy of time*. Oxford: Oxford University Press.

Price, H. & Weslake, B. (2009). The time-asymmetry of causation. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *The Oxford Handbook of Causation* (pp. 414–443). Oxford: Oxford University Press.

Psillos, S. (2002). *Causation & Explanation*. Stocksfield: Acumen.

Psillos, S. (2006). The structure, the whole structure and nothing but the structure? *Philosophy of Science*, 73, 560–570.

Psillos, S. (2007). Causal explanation and manipulation. In J. Persson & P. Ylikoski (Eds.), *Rethinking Explanation* (pp. 93–107). Springer.

Pust, J. (2012). Intuition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Winter 2012 edition.

Quine, W. V. (1969). Epistemology naturalized. In *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Rédei, M. (2002). Reichenbach's common cause principle and quantum correlations. In T. Placek & J. Butterfield (Eds.), *Non-locality and modality*. Dordrecht: Kluwer Academic.

Rédei, M. & Summers, S. J. (2007). Quantum probability theory. *Studies in the History and Philosophy of Modern Physics*, 38, 390–417.

Redhead, M. (1987). *Incompleteness, Nonlocality and Realism. A Prolegomenon to the Philosophy of Quantum Mechanics*. Oxford: Clarendon Press, 2. edition.

Redhead, M. L. G. (1982). Quantum Field Theory for Philosophers. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume 2: Symposia and Invited Papers (pp. 57–99).

Reed, M. & Simon, B. (1979). *Methods of Modern Mathematical Physics. III: Scattering Theory*. San Diego: Academic Press.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.

Reichenbach, H. (1959). The principle of causality and the possibility of its empirical confirmation. In *Modern Philosophy of Science*. London: Routledge & Kegan Paul.

Reutlinger, A. (2014). Can interventionists be neo-Russellians? Interventionism, the open systems argument, and the arrow of entropy. *International Studies in the Philosophy of Science*, 27(3), 273–293.

Rohrlich, F. (1999). On the ontology of QFT. In T. Y. Cao (Ed.), *Conceptual Foundations of Quantum Field Theory* (pp. 357–367). Cambridge: Cambridge University Press.

Rohrlich, F. (2000). Causality and the arrow of classical time. *Studies In History and Philosophy of Modern Physics*, 31(1), 1–13.

Ross, D. & Spurrett, D. (2007). Notions of cause: Russell's thesis revisited. *British Journal for the Philosophy of Science*, 58(1), 45–76.

Russell, B. (1912-1913). On the notion of cause. *Proceedings of the Aristotelian Society*, 13, 1–26.

Ryder, L. H. (2002). *Quantum Field Theory*. Cambridge: Cambridge University Press, 2. ed., repr. edition.

Sachs, R. (1987). *The physics of time reversal*. Chicago: University of Chicago Press.

Salart, D., Baas, A., Branciard, C., Gisin, N., & Zbinden, H. (2008). Testing the speed of 'spooky action at a distance'. *Nature*, 454.

Salmon, W. C. (1980a). Causality: Production and propagation. In *PSA: Proceedings of the biennal meeting of the Philosophy of Science Association*, volume II: Symposia and invited papers (pp. 49–69).

Salmon, W. C. (1980b). Probabilistic causality. *Pacific Philosophical Quarterly*, 61, 50–74.

Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton (New Jersey): Princeton University Press.

Salmon, W. C. (1989). Four decades of scientific explanation. In P. Kitcher & W. C. Salmon (Eds.), *Scientific Explanation*, volume XIII of *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press.

Salmon, W. C. (1994). Causality without counterfactuals. *Philosophy of Science*, 61(2), 297–312.

Salmon, W. C. (1997). Causality and explanation: A reply to two critiques. *Philosophy of Science*, 64(3), 461–477.

Salmon, W. C. (1998). *Causality and Explanation*. Oxford: Oxford University Press.

Savitt, S. F. (1996). The directions of time. *British Journal for the Phillosophy of Science*, 47(3), 347–370.

Schaffer, J. (2001). Review: Physical causation. *British Journal for the Phillosophy of Science*, 52, 809–813.

Schaffer, J. (2004). Causes need not be physically connected to their effects: The case for negative causation. In C. Hitchcock (Ed.), *Contemporary debates in philosophy of science*. Malden: Blackwell Publishing.

Schaffer, J. (2013). The metaphysics of causation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Winter 2013 edition.

Scheibe, E. (2007). *Die Philosophie der Physiker*. Verlag C. H. Beck.

Schlosshauer, M. (2004). Decoherence, the measurement problem, and interpretations of quantum mechanics. *Reviews of Modern Physics*, 76(4), 1267–1305.

Schlosshauer, M. (2007). *Decoherence and the quantum-to-classical transition*. Berlin: Springer.

Schulten, K. (2000). Notes on quantum mechanics.

Seevinck, M. P. (2010). Can quantum theory and special relativity peacefully coexist? arXiv: 1010.3714v1 [quant-ph].

Seevinck, M. P. & Uffink, J. (2011). Not throwing out the baby with the bathwater: Bell's condition of local causality mathematically 'sharp and clean'. In D. Dieks, W. Gonzalo, T. Uebel, S. Hartmann, & M. Weber (Eds.), *Explanation, Prediction, and Confirmation* (pp. 425–450). Springer.

Simon, C., Bužek, V., & Gisin, N. (2001). The no-signaling condition and quantum dynamics. *Physical Review Letters*, 87, 1–4.

Simon, H. A. & Rescher, N. (1966). Cause and counterfactual. *Philosophy of Science*, 33(4), 323–340.

Skyrms, B. (1984). EPR: Lessons for metaphysics. *Midwest Studies in Philosophy*, 9(1), 245–255.

Sperber, D., Premack, D., & Premack, A. (1995). *Causal Cognition*. Oxford: Oxford University Press.

Srednicki, M. (1993). Is purity eternal? *Nuclear Physics B*, 410, 143–154.

Stein, H. (1970). On the notion of field in Newton, Maxwell and beyond. In R. H. Stuewer (Ed.), *Historical and Philosophical Perspectives of Science*, volume 5 of *Minnesota Studies in the Philosophy of Science* (pp. 264–287). Minneapolis: University of Minnesota Press.

Sternberg, S. (1995). *Group Theory and Physics*. Cambridge: Cambridge University Press, paperback edition.

Suárez, M. (2004). Causal processes and propensities in quantum mechanics. *Theoria. Revista de Teoría, Historia y Fundamentos de la Ciencia*, 19(3), 271–300.

Suppes, P. (1970). *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.

't Hooft, G. (2007). The conceptual basis of quantum field theory. In J. Butterfield & J. Earman (Eds.), *Philosophy of Physics. Part A*, Handbook of the Philosophy of Science. Amsterdam: Elsevier B. V.

Teller, P. (1986). Relational holism and quantum mechanics. *British Journal for the Philosophy of Science*, 37, 71–81.

Teller, P. (1995). *An Interpretive Introduction to Quantum Field Theory*. Princeton (New Jersey): Princeton University Press.

Ticciati, R. (1999). *Quantum Field Theory for Mathematicians*. Cambridge: Cambridge University Press.

Unruh, W. G. & Wald, R. M. (1995). On evolution laws taking pure states to mixed states in quantum field theory. *Physical Review D*, 52(4), 2176–2182.

Vickers, P. (2013). *Understanding Inconsistent Science*. Oxford: Oxford University Press.

von Neumann, J. (1932). *Mathematische Grundlagen der Quantenmechanik*. Berlin: Verlag von Julius Springer.

von Wright, G. H. (1971). *Explanation and Understanding*. London: Routledge & Kegan Paul.

Wallace, D. (2008). Philosophy of quantum mechanics. In D. Rickles (Ed.), *The Ashgate Companion to Contemporary Philosophy of Physics* (pp. 16–98). Hants: Ashgate.

Wallace, D. (2014). Deflating the Aharonov-Bohm effect. arXiv: 1407.5073 [quant-ph].

Wallace, D. M. W. (2001). Emergence of particles from bosonic quantum field theory. arXiv: /quant-ph/0112149v1.

Wallace, D. M. W. (2011). Taking particle physics seriously: A critique of the algebraic approach to quantum field theory. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 42(2), 116–125.

Wayne, A. (2002). A naive view of the quantum field. In M. Kuhlmann, H. Lyre, & A. Wayne (Eds.), *Ontological Aspects of Quantum Field Theory* (pp. 127–134). New Jersey: World Scientific.

Weinberg, S. (1995). *The Quantum Theory of Fields*, volume I, Foundations. Cambridge: Cambridge University Press.

Weingard, R. (1982). Do virtual particles exist? *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1982, 235–242.

Weingard, R. (1988). Virtual particles and the interpretation of quantum field theory. In H. R. Brown & R. Harré (Eds.), *Philosophical Foundations of Quantum Field Theory* (pp. 43–58). Oxford: Clarendon Press.

Weinstein, S. (2006). Superluminal signaling and relativity. *Synthese*, 148(2), 381–399.

Wigner, E. (1939). On unitary representations of the inhomogeneous Lorentz group. *The Annals of Mathematics, Second Series*, 40, 149–204.

Wigner, E. (1963/1983). The problem of measurement. In J. A. Wheeler & W. H. Zurek (Eds.), *Quantum Theory and Measurement*. Princeton: Princeton University Press.

Williamson, J. (2006). Causal pluralism versus epistemic causality. *Philosophica*, 77, 69–96.

Williamson, T. (2007). *The Philosophy of Philosophy.* Malden: Blackwell Publishing.

Wohlfarth, D. (2013). *On the Conception of Fundamental Time Asymmetries in Physics.* PhD thesis, University of Bonn, Germany.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation.* Oxford: Oxford University Press.

Woodward, J. (2009). Agency and interventionist theories. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *The Oxford Handbook of Causation.* Oxford: Oxford University Press.

Wüthrich, A. (2012). Interpreting Feynman diagrams as visual models. *Spontaneus Generations: A Journal for the History and Philosophy of Science*, 6(1), 172–181.

Wüthrich, C. (2011). Can the world be shown to be indeterministic after all? In C. Beisbart & S. Hartmann (Eds.), *Probabilities in Physics.* Oxford: Oxford University Press.

Zee, A. (2003). *Quantum Field Theory in a Nutshell.* Princeton, Oxford: Princeton University Press.

Zeh, H. D. (2000). The meaning of decoherence. In P. Blanchard, D. Giulini, E. Joos, C. Kiefer, & J.-O. Stamatescu (Eds.), *Decoherence; Theretical, Experimental, and Conceptual Problems. Proceedings of a Workshop Held in Bielefeld, Germany, 10-14 November 1998.* Berlin, Heidelberg: Springer.

Zeh, H. D. (2006). Remarks on the compatibility of opposite arrows of time II. *Entropy*, 8(2), 44–49.

Zeh, H. D. (2007). *The Physical Basis of the Direction of Time.* Springer, 5th edition.

Zeh, H. D. (2012). Open questions regarding the arrow of time. In L. Mersini-Houghton & R. Vaas (Eds.), *The Arrows of Time.* Springer Verlag.

Zettili, N. (2009). *Quantum Mechanics: concepts and applications.* Chichester: Wiley, 2. edition.

Źukowski, M. & Brukner, Č. (2014). Quantum non-locality - it ain't necessarily so... *Journal of Physics A: Mathematical and Theoretical*, 47(42).

Zurek, W. H. (1982). Environment-induced superselection rules. *Physical Review D*, 26(8), 1862–1880.

Zurek, W. H. (2003). Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics*, 75, 715–775.