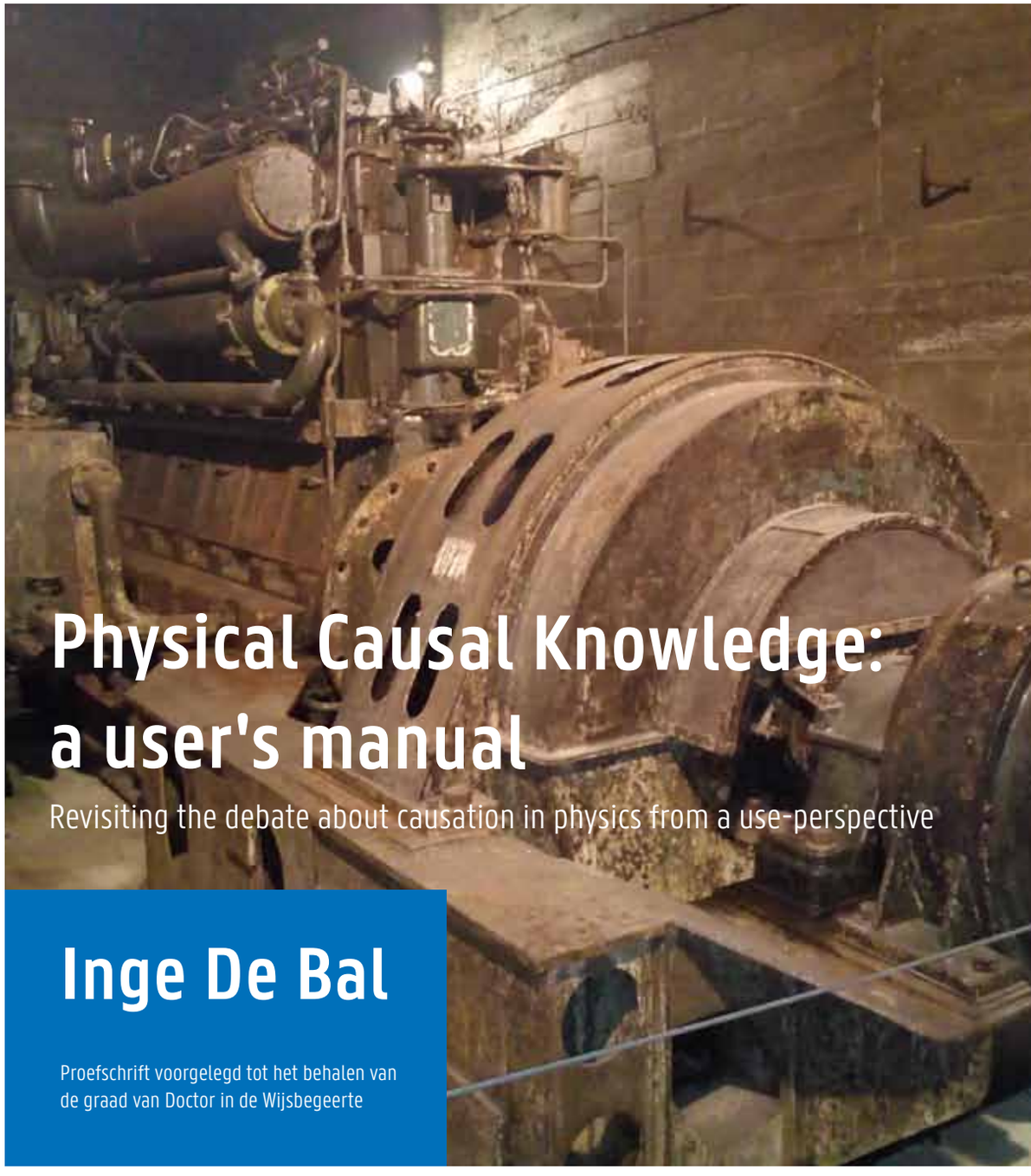
Physical Causal Knowledge: a user's manual Revisiting the debate about causation in physics from a use-perspective

Inge De Bal

FACULTEIT LETTEREN
EN WIJSBEGEERTE



Physical Causal Knowledge: a user's manual

Revisiting the debate about causation in physics from a use-perspective

Inge De Bal

Proefschrift voorgelegd tot het behalen van
de graad van Doctor in de Wijsbegeerte

UNIVERSITEIT
GENT

Promotor	Prof. dr. Erik Weber
Copromotor	Prof. dr. Phyllis Illari
Copromotor	Prof. dr. Bert Leuridan

Decaan	Prof. dr. Marc Boone
Rector	Prof. dr. Rik Van de Walle

FACULTEIT LETTEREN
EN WIJSBEGEERTE

Physical Causal Knowledge: a user's manual

Revisiting the debate about causation in physics from a use-perspective

Inge De Bal

Proefschrift voorgelegd tot het behalen van de graad van Doctor in de Wijsbegeerte
2017

UNIVERSITEIT
GENT

Acknowledgements

During the four years that I worked on this dissertation, I have received a lot of support from various people.

It goes without saying that my three supervisors have contributed enormously to both the quality and the existence of this dissertation. Each of them is due special thanks. First, I would like to thank Erik Weber, for introducing me to philosophy of science, for noticing me as a student and for encouraging me to pursue an academic career. If it had not been for him, I would not have started this dissertation. I would also like to thank him for allowing me to work on his project while I wrote my FWO-application, and for general guidance and feedback throughout these last four years. Bert Leuridan was also there from the start. I want to thank him for the thorough feedback that he remained to give even when he started his job in Antwerp. Finally, I am enormously grateful to Phyllis Illari, who found me hiding in Ghent and dragged me to the land of the coconut milk (more commonly known as London). Not only has she helped me formulate philosophical ideas I did not know I had, she introduced me to many like-minded philosophers of science and to the lovely philosophical community at STS UCL. This PhD would not have been here if it wasn't for her.

I am very grateful to my committee for agreeing to assess the result of these past four years. I hope it reads better than it wrote.

I thankful to the department of philosophy in Ghent for giving me several desks and cupboards while I worked on this PhD, and for providing caffeine when needed. I am also grateful to the members of the department. Thank you for sharing your passion and critical thoughts with me, for providing the general framework for my research experience and for broadening my fields of interests.

I also had the pleasure of spending some time in the PhD-room of the STS department at UCL. To the cohabitants of this room and Team Philosophy in general: for all the pitchers

of beer in that student bar I still can't find by myself, for dumpling nights and lip-syncing competitions, for wasabi lunches and coffee from the van on the street corner, thank you.

Thanks to Gitte for providing challenging quiz material, and for helping with the design of this book. I also want to thank Lut and Emilie for vital administrative support and non-academic talks in the hallway.

Dietlinde and Julie, thank you for sharing an office with me and for helping make the centre a more enjoyable, less cubicle like place. I have loved getting to know the two of you. Annelies M. is both colleague and friend, and during this PhD I have come to value this combination more than ever. She has been an enormous inspiration and provided a much needed safe space. I am a better person and researcher for knowing her.

I am grateful to my non-academic and non-philosophy friends, for valuing my crazy irrational side, and for providing balance. You know who you are.

To my parents, thank you for supporting me in pursuing this PhD and for installing me with a persistent mind that was able to see it through.

Finally, I want to thank my wonderful Mathieu. He has been an enormous support throughout the final and most important stages of this PhD. Not only has he proofread every chapter, he endured my panic, my exhaustion and my countless hours in the bathtub with love. I am not sure what I have done to deserve this, but I am immensely grateful for receiving it. From the bottom of my hart: thank you.

List of Abbreviations

ADT - Armstrong, Dretske and Tooley
$\mathbf{C}_a$ – causal claim #a
CFPT – Comparative Failure Process Tracing
CitS – Causality in the Sciences
$\mathbf{E}_a$ – example #a
E_{Fa} – example from failure analysis #a
$\mathbf{Exp}_a$ – explanation #a
MC – Mackie Cause
MDC – Machamer, Darden and Craver
MRL – Mill, Ramsey and Lewis
NCF – Negative Causal Factor
PCF – Positive Causal Factor
$\mathbf{S}_a$ – physical setup #a
SPSP – Society for the Philosophy of Scientific Practice
SU – Strong Unanimity
WU – Weak Unanimity

Table of Contents

Introduction

> *The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm. (Russell 1912, p.1)*

Philosophy of science aims to critically engage with scientific inquiry, to gain a better understanding of what science is, how science works and what conclusions can be drawn from scientific studies. If we look at the historical development of philosophy, it is obvious that science has been an object of philosophical investigation from the very beginning. Aristotle, for example, already reflected on scientific inquiry, the inductive-deductive method, and causation (Losee 2001, pp.5-11).

On closer inspection though, it becomes clear that the field now bears little resemblance to how it started. This is not surprising given the continuous development of science and the long history of the field. At the moment, "philosophy of science" captures a broad array of topics, including attempts to characterise and understand a (universal) scientific method, the quest for a demarcation criterion of science, insight in how scientific discovery works, accounts of justification and many more:

> The central target of philosophy of science is to understand science as a cognitive activity. Some of the central questions that have arisen and thoroughly been discussed are the following. What is the aim and method of science? What rules, if any, govern theory-change in science? How does evidence relate to theory? How do scientific theories relate to the world? How are concepts formed and how are they related to observation? What is the structure and content of major scientific concepts, such as causation, explanation, laws of nature, confirmation, theory, experiment, model, reduction and so on? (Psillos 2007, p.x)

Most of these topics are epistemological; they focus on how science provides us with knowledge and what the properties of this knowledge are. Other questions arise from a more metaphysical corner:

> The question on which this book will focus is, 'ought we to believe in the unobservable entities postulated by our best scientific theories?' […]. (Ladyman 2002, p.8)

This dissertation cannot engage with all these topics. It will deal with epistemological questions regarding causation, evidence and laws. These topics are interrelated, as will become clear, but if one of them is central for my purposes, it is causation.

Though this is not a novel topic (as mentioned, Aristotle already wrote on it), the last few decades have meant a boost in the diversity of approaches and scientific disciplines that have been investigated with regard to the causal claims they make, what these claims mean and what evidence is needed to support them. An exception is physics, where the main focus of reflection has been whether the laws express causal information at all. This question is often tackled from a very mathematical perspective and many philosophers (Bertrand Russell, John Norton, Huw Price, Donald Gillies,…) have concluded that the laws of physics are void of causal information – witness the quote in the beginning of this introduction. On closer inspection however, the debate is guided by some hidden assumptions that may be debatable. By challenging these assumptions, I will create room for a different analysis and correspondingly, a place for physical causal knowledge.

The debate assumes that the mathematical expressions of the laws of physics carry all their information, and that the laws exhaust the content of physics. Correspondingly, if there is no reference to causality there, there is no causal information in physics. In this dissertation, I expand the debate in a substantial way. Instead of looking at mathematically expressed laws, I will look at the ways we use physical causal knowledge. A focus on applications is notoriously absent from philosophy of physics, yet there is a very common-sense reason for including them in philosophical analysis:

> the emphasis on 'use' reminds us that we study the natural and social world for a reason, and we also conceptualize causal relations for a reason. (Illari and Russo 2014, p.238)

Regarding physics, philosophers tend to assume that applications follow rather straightforwardly from the laws.[1] In this dissertation, I build on existing philosophical analyses developed in light of the special sciences (viz. the social and biomedical sciences), to show that this view is mistaken. Moreover, when we thoroughly look at applications, the claim that physics contains no causal information turns out to be untenable, and a whole array of assumptions with it.

I will pay attention to three topics: the meaning of physical causal claims, the evidence we need to warrant these uses and the relation these claims have to the laws of physics. This is also the order reflected in the dissertation: chapter 2 deals with meaning, chapters 3 and 4 deal with evidence and chapter 5 directly deals with the relation to laws. My dissertation has four specific aims which fit into the three topics above, and two more generic, programmatic aims. As to the specific aims, I want to provide arguments for the following claims:

(I) To account for our successful creating, explaining, repairing, and maintaining of artefacts, we need a lot of specific physical causal information of the right kind, both of the artefact and of the physical and social context it functions in.

(II) The evidence needed to argue for physical causal claims extends beyond the laws of physics. We also use mechanistic evidence.

(III) The laws of physics are not the only source of general knowledge that is used to reach epistemic goals. Another important source is generalising local causal knowledge.

(IV) When looking at use, the importance of the distinction between laws and non-laws becomes significantly less prominent and the focus shifts to contextual goals. As such, the focus that philosophy of physics puts on theory and fundamental laws does not aid us in understanding how we use physical causal knowledge to achieve epistemic goals.

The more generic aims are the following:

(A) To show that using and producing physical causal knowledge is not a trivial affair.

[1] The work of Nancy Cartwright is a notable exception. She stressed the importance of a use-perspective even in her early work like (1979).

(B) To show there a several serious philosophical issues connected to useful physical causal knowledge.

The relation between these groups of aims is as follows: (I) till (IV) are positions with respects to some of the issues mentioned in (B). The way in which these aims are reached differs: (I) till (IV) each have a dedicated chapter (as mentioned above), while (A) and (B) are realised by the dissertation as a whole. In the conclusions of each chapter, I will systematically address how the chapter contributed to arguing for the specific and the generic aims. My conclusions will contradict several traditional views and assumptions of philosophy of physics.

The conclusion is not the only non-conformist aspect of this dissertation: I also use an uncommon methodology and uncommon cases. Regarding the methodology, I will use philosophical tools from the special sciences to analyse physical cases. Because of the prestige that physics has as a science (and correspondingly, the prestige that philosophy of physics has), it is often used as an example for the special sciences. So using philosophy of the special sciences as a guidance for analysing physical cases is not that common. But as will become clear throughout this dissertation, doing so opens a lot of perspectives.

Regarding cases, I will mainly focus on contexts associated with creating, explaining, repairing and maintaining artefacts. This is a practice that pertains to the engineering sciences and technology. In general, reflections on engineering and on technology are not seen as part of philosophy of physics; they make up their own domain. Moreover, there is some controversy whether engineering, or design, constitute sciences. Philosophers of technology have attempted to demarcate their field by arguing that it has different features than science (e.g. (Galle and Kroes 2014)).

The cases will also increase in complexity throughout the chapters. This allows me to reflect on the different aspects of using causal knowledge (viz. meaning, evidence and relation to laws) in a methodical way, and to show that each of these aspects involves specific challenges worthy of philosophical attention.

Because of these somewhat controversial choices, I will first spend some time framing them in the next chapter. I will first present more background information needed to understand my dissertation properly. This information will also allow me to motivate my choices of research topics. I will scrutinise the theory-centeredness of philosophy of physics and its emphasis on laws. By comparing this with the debates in philosophy of the social and of the biomedical sciences, I will argue in favour of a more practice-engaged philosophy of physics and physical knowledge. I will also spend some time sketching the status of the debate on causality in theoretical physics, by comparing John Norton's position with that of Mathias Frisch. I will also situate my dissertation in the

literature on causality and on artefacts. Finally, I will reflect on the case-studies that I use throughout this dissertation. The chapter ends with a more specific formulation of my research questions and a more elaborate overview of what can be expected of the next chapters.

As a final remark, I would like to note that this dissertation may contain errors and typos. I wanted to have a final read to correct some of those. Unfortunately, as it goes, time did not agree.

References

Cartwright, Nancy. 1979. Causal laws and effective strategies. *Nous*:419-437.

Galle, Per, and Peter Kroes. 2014. Science and design: Identical twins? *Design Studies* 35 (3):201-231.

Illari, Phyllis, and Federica Russo. 2014. *Causality: Philosophical Theory meets Scientific Practice*. 1 edition ed: Oxford University Press.

Ladyman, James. 2002. *Understanding philosophy of science*: Psychology Press.

Losee, John. 2001. *A Historical Introduction to the Philosophy of Science*. Fourth Edition ed. Oxford, New York: Oxford University Press.

Psillos, Stathis. 2007. *Philosophy of science AZ*: Edinburgh University Press.

Russell, Bertrand. 1912. On the notion of cause. *Proceedings of the Aristotelian society* 13:1-26.

Chapter 1 Why, how, and what? Framing this dissertation

This dissertation engages with topics from philosophy of science, philosophy of physics, philosophy of engineering, philosophy of technology and philosophy of causation. To see how my dissertation relates to these disciplines and how it might contribute to them, I will first identify some tendencies in philosophy of science and philosophy of physics specifically that make interaction with philosophy of engineering and of technology particularly difficult, namely a focus on theory and taking physics as a paradigm science. I will then sketch how a Philosophy of Science in Practice (PSP) approach can make a difference. This will set up the general framework in which my dissertation should be understood and will allow me to specify my central questions further: questions regarding our use of physical causal knowledge. To relate my questions and the thesis in general to the literature, I will then pay some attention to what a focus on *use* can do, to the variety of theories of causality, and to artefacts and the engineering sciences. This discussion should give the reader enough background information to understand what I want to achieve with this work, and equally, what I do not want to do. Let's start with the theory-focus of philosophy of science.

1.1 The focus of philosophy of science

1.1.1 What is science?

The brief characterisation in the introduction showed that the field of philosophy of science is vast and diverse. But like every study of a specific field, doing philosophy of science hinges on some (implicit) conception of the domain of study and of the reason

why it is interesting or worthwhile to investigate. Philosophers have argued that before a thorough study of science can begin, we need a definition of what science is and which practices are mistakenly seen as scientific.

> If we want to understand how science works, it seems that the first thing we need to do is work out what exactly we are trying to explain. Where does science begin and end? Which kinds of activity count as "science"? […] There is a lot of disagreement about what counts as science. (Godfrey-Smith 2009, p.2)

Schurz connects this to the difference in societal status between science and non-science:

> The *demarcation* problem is highly significant in society. In this context, this question consists of which of our ideas have a claim to the status of objective knowledge that should be taught in public educational institutions, as opposed to subjective opinion, political values, ideologies or religious convictions. (2013, p.2)

Unfortunately, what society sees as science, is not straightforward either. Nor is it constant: what disciplines are seen as scientific changes throughout history. For example, it used to be quite controversial to call disciplines like economics and psychology "science". Currently, there is some consensus about their scientific status (Godfrey-Smith 2009, p.3). But many other fields are still contested. Godfrey-Smith mentions archaeology and anthropology (Godfrey-Smith 2009, pp.2-3), but in light of my dissertation, engineering and design definitely deserve some mentioning. So the societal view of science does not seem to provide much guidance to decide what to study. Moreover, for many philosophers of science, there is a normative dimension to the definition of science. Not everything that society sees as science deserves the title.

1.1.2 The changing domain of philosophy of science

The development of philosophy of science in the last 50 years or so shows that the disagreement Godfrey-Smith refers to is very real and in flux. What philosophers of science consider to be interesting (scientific) questions can change through time. The philosophy of biology for example, only really took off half a century ago (Godfrey-Smith 2009, p.3).

However, some disciplines enjoy a high degree of consensus and have enjoyed it for some time. In particular, philosophers of science seem to agree that *physics* is a science worth studying:

> People often think of physics as the purest example of science. Certainly physics has a heroic history and a central role in the development of modern science. (Godfrey-Smith 2009, p.2)

Other than the agreement about physics, what makes up the domain of philosophy of science is not agreed upon and in constant evolution. Because of this, physics took up a central role for some time. But this has been changing, and other disciplines are receiving more attention from philosophers of science, like economics and biology.

This is related to another trend in the philosophy of science, namely specialisation. More and more philosophical questions and debates are tackled from and tailored to one specific branch of science. This has been noticed by many philosophers of science.

> The questions above [What is the aim and method of science? What makes science a rational activity? What rules, if any, govern theory-change in science? How does evidence relate to theory? How do scientific theories relate to the world?] did not change. But the answers that were considered to be legitimate did - the findings of the empirical sciences, as well as the history and practice of science, were allowed to have bearing on, perhaps even to determine, the answers to standard philosophical questions about science. In the 1980's, philosophers of science started to look more systematically into the micro-structure of individual sciences. The philosophies of the individual sciences have recently acquired a kind of unprecedented maturity and independence. (Psillos 2007, p.x)

> One distinguished between *general* philosophy of science, and philosophies of the *particular* sciences. The latter ones are concerned with special kinds of disciplines such as philosophy of physics, biology, psychology, the social sciences, of the humanities. The general philosophy of sciences discovers those components of knowledge that are more or less common to all scientific disciplines. (Schurz 2013, p.2)

Psillos stated this a decade ago, but Schurz shows that philosophers are still thinking about the impact of specialisation.

This specialisation also affects the philosophy of causality (a subdomain in the philosophy of science). While, as I mentioned, questions into the nature of causality have occupied philosophers of science from the very beginning, during the last decades philosophy of causality diverged in the study of causalities in different scientific disciplines. In "Causality in the Sciences" (*Causality in the Sciences* 2011), for example, the topics range from causes of evolution (biology) over causality in medicine (health sciences), to the error term in econometrics (economics). All of these questions are raised within the context of the specific disciplines. As a result, the debates no longer necessarily aim for one all-encompassing notion of causality. More and more

philosophers now allow for different complementary notions that are appropriate for a specific scientific field, like biology or the health sciences. I will get back to this below.

1.1.3 Physics as the paradigm of science

While the specialisation resulted in more profound attention to some neglected sciences, it initially entailed a lot of comparison between physics and the so-called special sciences.

> The original idea was to do "philosophies of the special sciences" which, roughly, includes all the sciences except for physics. (Allhoff 2015, p.5)

This comparison was not always in favour of the new fields. Other (and often newer) fields of science were expected to mimic physics, with the danger of losing their scientific status if they failed.

> […] there has traditionally been a strong tendency to think of physics, which appears to be strongly focused on discovering laws of nature, as the very paradigm of a science. Other fields of science get to count as sciences only to the extent that they emulate the methods, theory forms, and successes of physics. (Roberts 2004, p.152)

This is a testimony to the central position that physics still played and plays in the philosophy of science, even after specialisation. So even though disciplines like biology and the health sciences received more attention and were considered to be legitimate topics of research, the historical dominance of physics lingered. The focus on mathemisation and laws, for example, was one way in which the debate about which disciplines were truly scientific, was influenced:

> Many philosophers of science have viewed physics as *the* science *par excellence*. It is certainly true that physics, and astronomy in particular, was the first empirical science to be rendered in a mathematically precise form. […] It would not have been unreasonable, then, for philosophers to predict that all genuine branches of science would ultimately come to look like physics: a few simple laws of vast scope and power. […] In the twenty-first century, we have come to learn better. Chemistry and biology have certainly advanced to the stage of scientific maturity, and they look nothing like the model of a scientific system built upon a few simple laws. In fact, much of physics does not even look like this. (Hitchcock 2004b, p.10)

Some philosophers went further than merely seeing physics as the paradigm of science. They believed physics would eventually replace all other science. The following quote is

from an article from 1974, which was reprinted in 2007 as part of an anthology of the philosophy of science:

> A typical thesis of positivistic philosophy of science is that all true theories in the special sciences should reduce to physical theories in the long run. […] I think that many philosophers who accept reductivism do so primarily because they wish to endorse the generality of physics *vis à vis* the special sciences: roughly, the view that all events which fall under the laws of any science are physical events and hence fall under the laws of physics. (Fodor 2007, p.445)

Reducibility of other sciences to physics is another topic where philosophy of physics set the terms. Though reduction is not something that philosophers of the special sciences necessarily worry about, it still remains a point of debate:

> In the 1960s and 1970s, philosophical attention began to be afforded to individual sciences in a new way. Again, this is not to imply that no philosophical attention had been given to those sciences before these decades, but rather that new emphases were born. While being careful not to overstate the case, the emergence of philosophy of biology played a large role in this transition. It is probably safe to say that those interested in biology had begun to tire of attempts to subsume biology under physics and were highly motivated to show that they had their own (irreducible) research program. (Allhoff 2015, p.4)

The specialisation was accompanied by attempts of the individual sciences to free themselves from the physics-centred philosophical focus. I want to draw specific attention to attempts to overturn the emphasis on *theory* and *laws* that the dominance of philosophy of physics carried over to the debate on the special sciences.

In philosophy of physics, a major object of research is the laws of physics. Physical theory is seen as a set of those laws, and studying them is believed to answer our questions with regard to science.

> The field of physics includes many principles referred to as “laws”: Newton’s laws of motion, Snell’s law, the law of thermodynamics, and so on. It also includes many “equations” - the most important being named after Maxwell, Einstein and Schrödinger - that function in exactly the same way that the so-called laws do. While none of these laws are universally true - they all fail within one domain or other - physics is clearly in the business of looking for universal laws, and most physicists believe that there are laws “out there” to be discovered. (Hitchcock 2004a, p.149)

Because physics was seen as the paradigm science, the same focus was taken in studying the special sciences like biology and sociology. From about the 1970s to well in the

2000s, this was a particularly influential topic: whether the regularities of biology and of social science are *laws*.

> A related challenge to the scientific status of the social sciences claims that science aims at the discovery of *laws*, and that there can be no genuine laws of social science. In the third quarter of the twentieth century, especially, it was believed that laws were essential for both explanation and confirmation by evidence. (Hitchcock 2004b, p.16)

Both debates looked similar: philosophers analysed the regularities of biology and sociology respectively, investigated their properties and compared them to the properties that the laws of physics were supposed to have. In biology, the main contrast was drawn between the *contingent* regularities of biology, that could have been different if the evolutionary tape had been played again (Beatty, 1995), and the *necessary* laws of physics.[1] I will introduce the debate on laws in the social sciences via two papers, one by Kincaid (2004) and one by Roberts (2004), where they argue respectively for and against laws in the social sciences.

According to Roberts,

> [l]aws of nature are regularities that have certain features: they are global or universal, and robust, in the sense that they do not depend on contingent details of particular systems of objects, and they would not be upset by changes in the actual circumstances that are physically possible. (2004, p.166)

He sees physics as committed to discovering laws of nature (Roberts 2004, p.153). In his article, he argues that any laws of the social sciences would always need to be hedged laws, because they are never without exception and cannot be reformulated in strict probabilistic dependencies (Roberts 2004, p.159). But, he argues, there are no hedged laws since they would "be or entail hedged regularities, and there is no coherent concept of a hedged regularity". (Roberts 2004, pp.162-163) However, this need not be a problem. The social sciences do not need laws to allow for predictions and explanations.

> […] projectibility would amount to it being rationally justified to expect similar statistic patterns to prevail elsewhere. This would be useful for predictive purposes, and on many models of explanation, it would be explanatory as well. None of this requires anything be considered a law […]. (Roberts 2004, p.165)

[1] I will return to this debate in chapter 5. There, I also reflect on different philosophical definitions of "laws".

Not everyone agrees with this analysis. Kincaid, in the same book, defends the opposite thesis: the social sciences do have laws. He also refers to physics for a "quintessential example of a law of nature", namely Newton's law (Kincaid 2004, p.170). Kincaid draws different conclusions from studying the examples of physics. He argues that Newton's law of motion identifies a causal factor and that there are regularities in the social sciences that do something similar. (Kincaid 2004, p.170)

> Finding a definition that fit [sic] philosopher's or even scientist's intuitions about what we call laws need not tell us much about the practice of science. […] The point of a philosophical account should instead be to shed light on the practice of science - in this case, what role laws play in science. […] Rather, our project should be to get clear enough on how laws function in science to ask the question as to whether the social sciences can function that way as well. Then what role do laws play in science? Perhaps many, but above all, science produces laws to explain and reliably predict the phenomena. That is precisely what identifying causal factors allows us to do. (Kincaid 2004, p.172)

Kincaid clearly works in a different tradition than Roberts, has different goals and even draws different conclusions about the nature of laws from the same example. This tells us several things. For one, taking physics as an example is a widespread practice. And second, the comparison is not as informative or conclusive as often thought. Depending on which reflections on physics the authors use as their guide, the conclusions differ. Kincaid identifies Newton's law of motion as pointing to a causal factor. Yet the idea that the laws of physics express causal information is highly debated - I will get back to this in the next section. Kincaid also refers to Cartwright to argue that failure to be universal is not a convincing argument against laws in the social sciences (Kincaid 2004, p.171). Roberts sees this failure as part of his main argument, because in his view, laws are universal. At the very least, physics is not the unambiguous, enlightening example that it is often considered to be, and neither is philosophy of physics. This insight is part of the way the special sciences are freeing themselves from the longstanding focus on physics.

1.1.4 Theory versus practice

What the debate in the social sciences also illustrates is the central position that scientific theory (and with that, the concept of *law*) has received in the philosophy of science. For a long time, whether a discipline had laws, had impact on the status it was

ascribed, on whether it counted as a science and to what extent, on the way we should treat the claims provided by this discipline. Yet several philosophers have begun to argue that laws - and more generally, theory[2] - are not all there is to science. Some of these criticisms were formulated in the context of debates on laws. In the debate regarding laws of biology, Craver and Kaiser argue that we should stop investigating whether the biological regularities are laws, and instead focus on how they succeed in helping us with prediction and explanation (2013, p.127)[3]. Roberts's reframing of regularities in the social sciences in terms of prediction and explanation involves a similar shift in focus, as well as Kincaid's plea for refocusing the question in terms of the roles of laws. But also on a more general level, philosophers have argued for a move away from merely studying theory towards a practice-based view of science. The Society of the Philosophy of Science in Practice (hereafter SPSP) makes this one of its main aims. Here are some extracts from its Mission Statement[4]:

> Philosophy of science has traditionally focused on the relation between scientific theories and the world, at the risk of disregarding scientific practice. […] We advocate a philosophy of scientific practice, based on an analytic framework that takes into consideration theory, practice and the world simultaneously.

The SPSP actively encourages reflections on knowledge that show awareness for and attention to the way knowledge is *shaped* by its intended use:

> Practice consists of organized or regulated activities aimed at the achievement of certain goals. Therefore, the epistemology of practice must elucidate what kinds of activities are required in generating knowledge. Traditional debates in epistemology (concerning truth, fact, belief, certainty, observation, explanation, justification, evidence, etc.) may be re-framed with benefit in terms of activities.

One of the consequences of a practice-engaged approach to science is a focus on disciplines that used to be seen as mere 'application' of theoretical knowledge, like engineering, pharmacology, and design. They list some of their specific points of attention:

---

[2] I will treat "theory" as roughly a set of connected laws.

[3] Note that Craver and Kaiser are not the first philosophers of biomedical sciences to draw the attention away from laws. Bechtel and Abrahamsen, for example, argue that it is mechanisms and not laws that are crucial for explanations in biomedical sciences (2005).

[4] See http://www.philosophy-science-practice.org/en/mission-statement.

1. We are concerned with not only the acquisition and validation of knowledge, but its use. Our concern is not only about how pre-existing knowledge gets applied to practical ends, but also about how knowledge itself is fundamentally shaped by its intended use. We aim to build meaningful bridges between the philosophy of science and the newer fields of philosophy of technology and philosophy of medicine; we also hope to provide fresh perspectives for the latter fields.
2. We emphasize how human artifacts, such as conceptual models and laboratory instruments, mediate between theories and the world. We seek to elucidate the role that these artifacts play in the shaping of scientific practice.
3. Our view of scientific practice must not be distorted by lopsided attention to certain areas of science. The traditional focus on fundamental physics, as well as the more recent focus on certain areas of biology, will be supplemented by attention to other fields such as economics and other social/human sciences, the engineering sciences, and the medical sciences, as well as relatively neglected areas within biology, physics, and other physical sciences.
4. In our methodology, it is crucial to have a productive interaction between philosophical reasoning and a study of actual scientific practices, past and present. This provides a strong rationale for history-and-philosophy of science as an integrated discipline, and also for inviting the participation of practicing scientists, engineers and policymakers.

A PSP approach not only involves studying more than what is written in scientific textbooks, but also studying other places and practices where knowledge gathering or knowledge construction takes place or is influenced. For example, research and development labs, where the research into new phenomena and the technical implementation of those phenomena go hand in hand (Freeman and Soete 2009, p.586); databases that are used to store and rank data; or even as Shapin (2016) noticed, the "food laboratories" of McDonalds, where the taste and ingredients of the Royal Cheese are constantly improved. These new contexts and practices give rise to new questions, and new answers. My dissertation will specifically connect with the first point of attention, namely the connection between knowledge and use.

Despite the noticeable growth in attention to more practice-engaged questions, not everyone is ready to 'jump aboard the practice train'. The focus on theory has a long history, and reframing philosophical debates towards scientific practice has led to a different conception of science. This is not an easy thing to do, and has significant consequences for the methodology of philosophy of science.

This is especially the case in philosophy of physics. I will sketch an episode in the discussion whether the laws of physics express causal information or are causal. This will show how a focus on scientific theory (and laws specifically) versus a focus on scientific practice can lead to different results and shape a debate. Specifically, I will compare John Norton's arguments with Mathias Frisch's and show how the latter attempts to broaden the philosophy of physics to include more practice.

1.2 Expanding philosophy of physics

1.2.1 Norton and Frisch on causality in physics

Whether the laws of physics express causal information is a longstanding topic in philosophy of science. Starting in the last decades of the nineteenth century, philosophers like Ernst Mach (1901), Bertrand Russell (1912) and John Norton (2007) have questioned whether causality has a place in physics. Their arguments are based on the mathematical formulation of the majority of the laws of physics, namely equivalencies.

Norton (2007) argues that mature sciences, "are adequate to account for their realms without need of supplement by causal notions and principles" (p.12). Causal notions belong to "earlier efforts to understand our natural world", or are helpful tools to explain science to laymen (ibid). According to Norton, if conforming a science to cause and effect makes no difference for the factual content of the science, then a notion of cause is empty and adds unnecessary baggage (2007, pp.3-4). So for "cause" to be relevant for science, it needs to influence the factual content of science in a significant and irreplaceable way. One common way that causation was thought to restrict the contents of our scientific theories, was via determinism:

> fix the present conditions sufficiently expansively and the future course is thereby fixed. (Norton 2007, p.4)

If that were the case, if the laws of physics allowed us to fix the future based on the present, then this would imply that some notion of cause was inherent in the laws. Norton argues that this is not the case. A deterministic notion fails altogether, since

quantum physics shows a failure of determinism, and Norton's famous thought-experiment about a frictionless dome shows that Newtonian systems can be indeterministic[5]. Moreover, a probabilistic notion of cause fails as well, since the quantum theories do not give us any specific probabilities regarding which future is more likely. Neither a deterministic view, nor even a probabilistic one hold up in light of our current mature theories. This is only one of the specific arguments Norton presents against a notion of cause in physics; he also developed more specific arguments tailored to e.g. modern physics (Norton 2006) and classical electrodynamics (Norton 2009).

I do not wish to go into these specific discussions, but I want to reflect on Norton's method for arguing against causes in physics. His method is based on the *theory* of physics: he analyses the laws and equations in physics without taking broader practices (modelling of actual phenomena, evidential practices,…) into account. This is especially clear from his account of folk causation. This notion is meant to capture the causal talk in scientific practice (Norton 2007, p.21). Yet this is not considered part of the mature sciences: the causal talk is merely the result of 'labelling' a specific property as causal because "we perceive some sort of commonality with a broader, vague notion of causality" (Norton 2006, p.231). Norton mainly contrasts this causal talk with a fundamental robust causal principle underlying all our sciences (2006, p.232). But what is also clear from his writings, is that this causal talk is not to be seen as a fundamental or sensible part of physics. Whether there is some causal principle or notion in physics can be determined by investigating the interplay of *the mathematical equations that constitute the laws of physics*. For example, when discussing the problems quantum mechanics raised for causality, he refers to what is allowed by the *theory*:

> Quantum theory brought other, profound difficulties for causation. Through its non-separability, quantum theory allows that two particles that once interacted may remain entangled, even though they might travel light years away from each other, so that the behavior of one might still be affected instantly by that of the other. (Norton 2007, p.5)

[5] In (2007) and (2008), Norton describes a mass placed on the top of a frictionless dome, in a gravitational field. "After remaining motionless for an arbitrary time, it spontaneously moves in an arbitrary direction, with these indeterministic motions compatible with Newtonian mechanics" (Norton 2008, pp. 786-787). Norton argues for this compatibility via the equation of movement of the mass, which has several solutions: one where the mass remains at the top forever, and one family of solutions that represents "spontaneous motion at an arbitrary time T in an arbitrary radial direction" (2008, p. 788). According to Norton "[t]he dome manifests indeterminism in the standard sense that a single past can be followed by many futures" (2008, p. 788).

Similarly, when discussing the indeterminacy of Newtonian mechanics:

> An important feature of Newtonian mechanics is that it is time reversible, or at least that the dynamics of gravitational systems invoked here are time reversible. This means that we can take any motion allowed by Newton's theory and generate another just by imagining that motion run in reverse in time. (Norton 2007, p.11)

This again shows that Norton bases his conclusions on the theory of physics.

As it goes in philosophical debates, not everyone agrees with Norton's analysis or conclusions. One of the contemporary defenders of a more substantial role for causes in physics, is Mathias Frisch. Frisch is not arguing in favour of the fundamental robust causal principle that Norton rejects; his focus is on revaluing the causal reasoning in physics:

> Thus, instead of asking for the metaphysical underpinnings of causal notions, the functional project asks what role, if any, causal notions play as part of our epistemic toolkit and as part of the representational resources. The legitimacy of causal notions or causal thinking is evaluated with respect to whether they serve a useful function, and any account of causation has to be defended with reference to the functional role of causal concepts. (Frisch 2014, p.9)[6]

While Norton dismisses causal talk, even by scientists, Frisch argues that positing causal structures is a fundamental part of physical scientific practice, without which "many of the inferences we routinely make in physics would simply be impossible" (Frisch p.21). To defend the importance of causal reasoning, Frisch first expands what philosophers like Norton implicitly consider to be the content of physics. Frisch starts from one of the basic epistemic goals of science: *representing* the world (2014, p.21). These representations can then be used for other epistemic goals, like prediction and explanation. When trying to understand how physics represents the world, Frisch posits that arguments[7] like Norton's offer little help. Because of their theory-focus, they are based on a faulty view of what science and specifically physics entails. Frisch frames it in terms of Cartwright's "vending machine view of science" (1999). The vending machine view is Cartwright's label for the idea that all you need to do in physics is give some input to the theory (specifically initial values) and after a bit of beeping, a representation or a

[6] Chapter 3 will show that it is not always clear what the boundaries between a metaphysical and a functional project are.

[7] To be precise, Frisch argues that many philosophers in this debate have used similar instances of the faulty argument scheme.

model will pop out, like a soda-can from a vending machine (Cartwright 1999, p.184). According to Cartwright this view is not compatible with how science works, and Frisch agrees.

Contrary to what philosophers who hold a vending-machine view think, fundamental equations and their mathematical consequences can never give us a complete understanding of how the world is represented via (physical) theory or laws. To supplement this view of science, Frisch argues that we need to take into account the user and the context, even for physical theories. His alternative account of scientific practice in physics starts from a "pragmatic and structural account of representation" (2014, p.37). The structural part is the claim that

> Representation is purely structural, since the models or representations employed, at least in the physical sciences, are mathematical structures and the only relevant resemblance between mathematical structures and physical systems is structural resemblance. (Frisch 2014, pp.27-28)

The similarity between mathematical structures and physical systems is structural in nature. The pragmatic part builds on Van Fraassen and is the claim that

> there is no representation except in the sense that some things are used, made, or taken, to represent things as thus and so. (Van Fraassen 2008, p.23)

Frisch argues that this pragmatic part solves certain problems associated with a purely structuralist account, such as the problem of explanatory asymmetry. In a pragmatic account of representation, there can be

> no "natural representations" – no naturally produced objects or phenomena that represent other phenomena without being used by someone to represent. (Frisch 2014, p.37)

In other words, we cannot simply let mathematical equations 'talk for themselves' in representing phenomena: the user and the context matter and cannot be ignored. Especially this latter part is fairly controversial, even though it explicitly builds on work in general philosophy of science by among others Cartwright (1983, 1999) and Van Fraassen (2008). Cartwright's books were published over fifteen years ago, but thinkers like Norton still mainly study the fundamental equations in isolation, which are seen as covering the entire content of a theory.

Following Frisch, once you take representations to be a serious part of physics, then you cannot ignore the role of causal reasoning anymore. For example, much physical modelling is supposedly done by solving a so-called initial value problem:

> Reasoning and inferences in physics can be exhaustively characterized in terms of a theory's dynamical models together with choices of particular initial and boundary conditions. (Frisch 2014, p.125)

The claim expressed here is a core-assumption of Norton (2007), but also of Huw Price and Brad Weslake (2009). The basic idea behind this assumption is that to model a specific system, the theory gives you the dynamical equations that apply to the system and you combine these equations with the initial values (e.g. the current state) of the system, to attain a model of the (future) states of the system (Frisch 2014, p.21). These dynamical models provide all the information we need and are attained in a mathematical way. Causal considerations are absent and irrelevant. Frisch counters this argument by showing that the choice between different possible dynamical models is in practice underdetermined and we can only choose between them by means of causal assumptions (Frisch 2014, p.125). I will get back to this in chapter 3. So while in theory, the equations might be all we need, in practice, we cannot get rid of causal reasoning without losing a significant part of scientific strength.

This comparison between Norton's and Frisch's methodologies for investigating the role of causes in physics, shows how much can change depending on how the boundaries of "science" are drawn. Specifically, taking the researcher, the modelling practices and the context into account, can determine whether there is a place for causes in physics, or there is not. Building on Frisch, I will also attempt to expand the subject of philosophy of physics by focusing on how we use physical knowledge in a causal way.

1.2.2 *Using* physical causal knowledge: explanation and intervention

Physics is the science par excellence when it comes to idealisation. In his analysis, Frisch explicitly moves away from looking at the equations in a conceptual vacuum. This is an enormous step towards a more practice-based philosophy of physics. By looking at modelling practices, he is focusing not only on the equations, but also on the way these equations are used to fulfil the epistemic goals we want to achieve with science. If we follow through on this broadening of physics, on this attention for the scientific practice and for how we succeed in reaching epistemic goals via science, a whole new array of interesting questions and contexts arises. Frisch focuses on representation. But science, including physics, also allows us to explain and predict phenomena, to intervene in the world and to manipulate it. One of the goals of science is to produce knowledge that we can *use*. Causal knowledge is used constantly throughout our lives, both in scientific contexts and everyday situations. In this dissertation, I will show that studying these

applications and the knowledge that is required to achieve them, sheds new light on the topic of physical causation, and on the role of laws for that matter.

First, I discuss some very general ways in which physical causal knowledge is used: explanation and intervention. To understand that physical causal knowledge allows us to *explain* physical phenomena, let's look at the well-known example of the flagpole. Suppose we have a flagpole that casts a shadow. The laws of geometrical optics (more precisely the law that $^{h}/_{l} = \tan\alpha$, where h refers to the height of the flagpole, l refers to the length of the shadow and α is the angle of elevation of the sun above the horizon) allow derivations in many directions. If we know the length of the shadow and the angle of the sun, we can calculate the height of the flagpole. But in the same way, we can calculate the length of the shadow by means of the height of the flagpole and the position of the sun. We can even determine the position of the sun by means of the height of the flagpole and the length of shadow. This is the symmetry mentioned above. These three derivations are not necessarily equally good as explanations, though. I assume that most readers will share the following judgements:

(Exp$_1$): Explaining the length of the shadow by means of the height of the pole and the position of the sun is a good causal explanation.

(Exp$_2$): Explaining the height of the pole by means of the length of the shadow and the position of the sun is a bad causal explanation.

(Exp$_3$): Explaining the position of the sun by means of the height of the pole and the length of the shadow is a bad causal explanation.

These judgements are based on the following causal belief:

(C$_1$): The position of the sun and height of the pole are causally relevant for the length of the shadow, but not the other way around.

If you believe instead that the position of the sun and the length of the shadow are causally relevant for the height of the flagpole (which might seem weird, but in some cases might be justified, I will get back to this in chapter 3), you have a different causal belief and correspondingly, a possibly different judgement about which explanation to prefer. Causal beliefs and knowledge are used to decide between different explanations.

Physical causal knowledge is also relevant to guide our actions and interventions; another way of using causal knowledge. We do things to arrive at certain effects: we use a pump to put more air in our tires and increase their internal pressure; we flip a switch to turn on a light; we use a crowbar to exercise greater force and open a crate. The knowledge on which we base these interventions is physical, and whether we believe these actions are useful or justified, depends on what beliefs we have. Consider the

example of pumping air into your tires. This is a goal-directed action: you want to increase the internal pressure of the tires. This aim or goal in itself does not guarantee a successful action. There is also a causal belief involved:

(C_2): Pumping air into my tires increases their internal pressure.

If you do not believe this, your action does not make sense. And correspondingly, if your belief turns out to be false, you will not pump air into your tires next time you have a flat one. You will perform another action that corresponds to the causal belief that you have at that point in time.

1.2.3 First characterisation of my research topics

So a very common sense look at the world around us shows the importance of physical causal knowledge. Physical causal knowledge is everywhere and more importantly, it is *used* everywhere. It is used for explanations of physical phenomena, it is used for interventions in the natural world, and it is used when engaging with artefacts. This is already clear from my tire example: we use a pump in order to inflate our tires. The pump (viz. an artefact) is used to achieve a certain effect (viz. the repair of another artefact). This is an everyday example and may not look very interesting. But engineers and technicians do exactly that: they use physical causal knowledge to research, design, maintain and repair artefacts.

Regardless of the omnipresence of physical causal applications in the world, philosophers like Norton have explicitly argued against the need for causal notions in science. Norton did allow causal notions to play a role in certain contexts, but these causal notions are part of what he calls “folk science”:

> a crude and poorly grounded imitation of more developed sciences. (Norton 2007, p.2)

He compares this to specific situations where we still use Newton’s theory of gravitation, instead of Einstein’s theory of general relativity: it can be useful in certain situations. According to Norton, in many familiar contexts it is just “conceptually easier and quite adequate to imagine that gravity is a force or heat a fluid” (Norton 2007, p.13). Similarly, there are many familiar contexts in which we can describe physical processes in a causal way, using this folk science. So Norton seems to grant this folk science some authority.

In general, however, it seems that Norton does not believe the folk science to be worthy of much philosophical reflection. For instance, ascribing causal relations to phenomena is apparently straightforward:

> How do we know which terms in the science to associate with the cause and effect? There is no general principle. In practice, however, we have little trouble identifying when some process in science has the relevant productive character that warrants the association. Forces cause the effect of acceleration; or heat causes the effect of thermal expansion; or temperature differences cause the motion of heat by conduction; […]. The terms in the causal relation may be states at a moment of time; or entities; or properties of entities. (Norton 2007, p.16)

For the examples he mentions, it might be easy to determine the causal relation and correspondingly, to successfully ascribe[8] a causal relation to the phenomenon. However, in this dissertation, I will argue that when we study real circumstances where we *use* and produce physical causal knowledge (be it in relatively easy and day-to-day circumstances or more uncommon complex ones) the situation is way less clear.

Norton is not the only philosophers that takes this position. Because of the traditional focus on laws and theory, disciplines like engineering have been ignored in philosophy of science, since there is no 'fundamental theory of engineering'. After all, engineering was long considered to be nothing but an application of the all-encompassing laws of more fundamental sciences, like physics. As a consequence, it was believed that studying engineering practice would not teach us anything we couldn't get from the laws of physics. Philosophers like Cartwright, Van Fraassen and Frisch have started to break down the idea that (fundamental) laws are the only things worth studying in physics. Philosophers in the SPSP have argued the same in a more general way. In this dissertation, I contribute to their project.

To do so, I will study this causal knowledge from the *use*-perspective and show that, pace Norton, it is worthy of philosophical reflection. With the "use-perspective', I refer to my focus on the way we use and produce physical causal knowledge. Specifically, I will investigate the physical (causal) knowledge necessary for and acquired in researching, building, using and maintaining artefacts[9], with regard to

(1) the meaning of causal claims about artefacts,

(2) the evidence needed for using this physical causal knowledge, and

(3) the relation of this physical causal knowledge to the laws of physics.

[8] I will use "ascriptions of causal relations" as synonymous with "making causal claims".

[9] Though I mainly limit the use-contexts to artefacts, my analysis also holds for natural contexts. I will therefore on occasion explain how it can be adapted.

I will show that all these topics related to using physical causal knowledge raise interesting philosophical issues, which will constitute an argument for my two more generic aims described in the Introduction (viz. (A) and (B)). With regards to the first topic, I will for instance investigate what we mean when we say that X is a cause of Y in the case of artefacts. Problems regarding evidence, the second topic, are for instance related to the specific evidence that we need to use physical causal knowledge for explanations of or interventions in artefacts. The third topic is pretty straightforward and deals with the debate I discussed in 1.2.1. One question to answer is for instance: If the laws of physics do not contain causal information, then how is our physical causal knowledge related to them? These topics are related to my four specific claims that I described in the Introduction. I already presented the background information to understand the third topic. However, to ensure that my first two questions are clear, I will discuss relevant parts of the literature on causality (1.3) and on evidence for use (1.4).

1.3 Causality - what I am and am not talking about

Focusing on explanation and intervention intuitively opens the conceptual space for studying physical causal knowledge. Besides Norton and Frisch, many other philosophers have made some contribution to the literature on causation in science. I mentioned in the Introduction that Aristotle already wrote about causality, and it has remained a popular topic since. The specialisation I mentioned in the introduction sparked a diversification of accounts. Because so much has been written, it would be ridiculous to attempt a full overview of the literature. Moreover, plenty of other philosophers have written excellent works on the topic (for example (Psillos 2002), (*The Oxford handbook of causation* 2009), (*Causality in the Sciences* 2011), (Illari and Russo 2014)). What I mainly want to do is explain how my dissertation relates to other contributions and projects relating to causality.

First, it used to be quite common to present *accounts* or theories of causation: characterisations of what causation is. For example, Cartwright presents this list of theories of causality:

- the probabilistic theory of causality (Patrick Suppes) and its descendants
  - Bayes-nets theories (Wolfgang Spohn, Judea Pearl, Clark Glymour);
  - Granger causality (economist Clive Granger);

- modularity accounts (Pearl, James Woodward, economist Stephen LeRoy);
- manipulation accounts (Peter Menzies, Huw Price);
- invariance accounts (Woodward, economist/philosopher Kevin Hoover, economist David Hendry);
- natural experiments (Herbert Simon, economist James Hamilton);
- causal process theories (Wesley Salmon, Philip Dowe);
- the efficacy account (Kevin Hoover);
- counterfactual accounts (David Lewis, Hendry, social scientists Paul Holland and Donald Rubin). (2007, p.43)

In the context of physics, Phil Dowe's conserved quantity theory is probably most famous. In Dowe's theory, causes are processes, and specifically those processes in which a conserved quantity is exchanged (Dowe 2000, p.89).[10] In the biomedical sciences on the other hand, Woodward's interventionist theory (2003) is currently very popular. This reflects the specialisation I talked about above. Different theories have been developed in the context of different sciences (especially biomedical and social sciences), to accommodate the different methods and claims.

1.3.1 Epistemology vs. metaphysics

A first very general way of distinguishing between these accounts of causation is between epistemological projects and metaphysical ones. Metaphysical projects study causation from an ontological perspective, such as what Russell (1912) called "the law of causality". Norton summarised the law of causality as follows:

> that every effect is produced through lawful necessity by a cause (2007, p.11).

More generally, metaphysics of causation answers questions like 'What are the causal relata?', 'What characterises them?', 'How many relata are there?' (Schaffer 2016).

Epistemological projects on the other hand, attempt to get insight into causal knowledge, how we get it, what properties it has etc. My dissertation clearly belongs to the epistemological project, since I am explicitly studying causal knowledge.

Note that it is not always clear to which project a causality account belongs. Though epistemological and metaphysical projects attempt to answer different questions, they

[10] Conserved quantities are quantities that follow a conservation law, described by physics. Examples are mass-energy, linear momentum and charge.

do not have to be completely distinct enterprises. On the contrary, philosophers like Cartwright (2007) or Illari and Russo (2014), have argued that they should go hand in hand. Regarding causality, a methodological approach that allows both metaphysical and epistemological questions is the Causality in the Sciences (CitS) Approach[11], as found in (*Causality in the Sciences* 2011) and described by (Illari and Russo 2014, pp.201-210). It is characterised by a close and often simultaneous engagement with both philosophical and scientific literature, in the hope of contributing to scientific practice and policy.

> We suggest instead successful philosophy of causality is iterative, moving freely between science and philosophy, neither first, often simultaneously studying both literatures. It proceeds, ideally, in a dialogue between philosophy and science—which is sometimes, but not always, a dialogue between scientists and philosophers! (Illari and Russo 2014, p.207)

Science and scientific literature are used to generate new ideas, select problems and test accounts (Illari and Russo 2014, pp.208-209). My dissertation explicitly engages with scientific practice and literature. Because of this, it fits in the CitS approach. This attitude has been present during my research and will resonate throughout this dissertation. The focus will be epistemological questions, but where relevant I will draw related metaphysical conclusions. [12]

1.3.2 Token versus type

Another distinction is between token- and type-level theories. Token-level theories capture causation between *events*, while type-level theories capture causation between *populations* or between *types*. In token-level theories, we distinguish between process theories (e.g. (Dowe 2000), (Salmon 1998)) and difference-making theories. Difference-making theories come in three varieties: regularity theories (e.g. the INUS account of (Mackie 1980)), probabilistic theories (e.g. (Suppes 1970), (Glymour 2001)) and counterfactual theories (e.g. (Lewis 1973)). All of these theories have their merits and problems. Process theories have a notorious difficulty dealing with negative causation (prevention and causation by absence). Processes cannot involve absent objects. So they

[11] As a testimony to the diversity of questions that can be tackled from the CitS-perspective, see the range of topics in (*Causality in the Sciences* 2011). The contributions of Stuart Glennan (2011), Phil Dowe (2011) and Illari & Williamson's (2011) are for instance more focused on metaphysical questions. The contributions by e.g. Gillies (2011), Harold Kincaid (2011) or Bert Leuridan & Erik Weber (2011) tackle epistemological questions.

[12] Chapter 5, for example, will deal with the consequences of my project for the metaphysics of laws.

cannot account for the lack of water causing my plants to die. Difference making theories do not have this problem. They basically focus on statements of the form "Occurrence/absence of event c causes occurrence/absence of event e". However, they cannot express the quantitative side of causal relations (e.g. how much pressure increase due to how much temperature increase).

Type-level theories include the interventionist account of Woodward (2003), Bayes-net theories like (Pearl 2009), the independent alterability account (Hausman 1998), the comparative model (Giere 1997) and the probabilistic model (Eells 1991). All these theories come with a specific set of assumptions: Giere requires the acceptance of hypothetical populations, Pearl requires a lot of mathematical modelling, Woodward has been criticised for needing a causal concept (namely intervention) to define causation. So far, no one theory has been able to win over everyone else. In 1.3.3., I will reflect on the possibility of finding such one theory.

What has also puzzled philosophers is the relation between the two. Some philosophers have argued that we need two concepts of causation, like Elliott Sober (1984), one token level and one type. Elsewhere, Sober (1986) argued that the two are connected with regard to evidence as well:

> […] the more evidence we have about the effects of smoking on lung cancer in general, that is, the more evidence for the generic claim, the more likely (ceteris paribus) the single-case relation will be, that is, that smoking causes lung cancer in individual patients. (Illari and Russo 2014, p.42)

Moreover, there is the metaphysical question of which level is ontologically primary to the other. Mackie and Cartwright argued that the single-case is primary, but there is no clear agreement in the literature (Illari and Russo 2014, p.44).

The majority of this dissertation engages with type level causal claims, but this does not reflect the conviction that type level is primary or more important. Most of my cases happen to deal with type level claims. But, as will become clear throughout the chapters, I also engage with token level claims and the relation between token and type level.

1.3.3 Monism vs. pluralism

A third separation in the literature is the one between causal monists and causal pluralists. Monists hold that one theory suffices to capture causality. This can be combined with the metaphysical view that there is only one type of causal relation in the world. As Illari and Russo argue, most theories that Cartwright summarises were intended monistically, but are not considered successful in that respect (Illari and Russo 2014, p.249). As they argue, pluralism offers an alternative. But pluralism, like cause for

that matter, can mean many things. Illari and Russo argue that one can be a pluralist regarding (1) different types of causing in the world (e.g. pushing and pulling convey different information about how the cause affects the effect); (2) different concepts of causation (e.g. depending on scientific discipline); (3) different types of inferences (e.g. those related to prediction can differ from those related to explanation) ; (4) different sources of evidence for causal relations (e.g. mechanistic and correlational); (5) different causal methods (e.g. Bayesian nets and structural models) (Illari and Russo 2014, pp.250-254). The pluralism they advocate combines all of these, in the sense that "the full array of concepts of causality developed within philosophy could very well be of interest to practising scientists, if explained in terms of the scientific problems most familiar to them" (Illari and Russo 2014, p.256). With "the full array of concepts", they refer to concepts developed regarding all five points that can be looked at pluralistically, viz. causal relation, causal concepts, causal inferences, causal evidence, causal methods.

My dissertation fits in this broad pluralism. I do not claim that my analyses hold for all causal relations, not even for all causal relations or inferences in engineering sciences. I do claim that they can be useful for scientists, when aptly reformulated when necessary. As such, I believe that this dissertation can be a useful part of what Illari and Russo refer to as "the library of concepts" that can aid scientists in their actions (Illari and Russo 2014, p.256).

1.3.4 Combining the three choices

So my take on causality can be situated in a pluralistic epistemological project. Correspondingly, I do not take stance in the debate whether token-level is primary to type-level. I take causal theories to be tools to build models, and depending on the context, a token or type-level notion of causality may be necessary. I also do not believe one level is "more correct" than another in the absence of specific cases. On the contrary, in chapter 2, I will use a combination of different theories to elucidate my cases. Since I am trying to understand and represent physical causal knowledge and am committed to a pluralistic notion of causality, this is not a problem.

1.4 Evidence for *use* in the biomedical and social sciences

Philosophy of the social and biomedical sciences already focuses a lot on evidence for use and the applicability of causal knowledge. Looking at the sciences they study, there are good reasons for doing so. That a policy action or a medicine works in one context, does not mean it will certainly work in another. Consider an example by Cartwright and Hardie (2012, pp.80-84) about policy actions directed at improving the nutritional status of pregnant women and their children in Bangladesh. The actions consisted of providing counselling for pregnant women and giving supplementary feeding for new-borns. It was modelled after a successful policy action in Tamil Nadu. However, in Bangladesh, the actions had no noticeable impact on nutritional standards. The reason was that in Bangladesh the mothers were not responsible for the shopping (so the counselling did not lead to better choices in food purchasing) and the supplementary food was often passed down to other members of the family and used as substitute (so the children did not actually eat the supplementary foods) (Cartwright and Hardie 2012, p.82). This is often referred to as the problem of extrapolation or external validity:[13]

> Evidence is always collected in some population in some circumstances. With most methods the inferences that are licensed from that method are tied to the populations and situations in which the evidence is obtained and licence to go beyond those must come from somewhere outside that method. (Cartwright and Hardie 2012, pp.38-39)

Nancy Cartwright was one of the first philosophers to draw attention to the connection between knowledge and use. It was already present in her first paper on causation (viz. (Cartwright 1979)) and remained important in the rest of her work. In *Hunting Causes and Using Them* (2007), she argues that a theory of causation should both tell us how to hunt causes and how to use them. If we do not take use into account, we get our analysis of causation wrong because we miss the essential feature. According to her, many theories of causation fail on this criterion:

> They provide an elaborate procedure for deciding when we can attach the label 'cause'. But then what? There is nothing more in the account that allows us to move anywhere from that, nothing that licenses any inferences for use. (Cartwright 2007, p.49)

[13] I will return to the debate on extrapolation in chapter 3.

Cartwright's initial focus was on physics, but she went on to mainly study the social sciences, and in particular, policy. She remained to name policy and technology in the same breath, for example when discussing the gap between Woodward's account of causation and contexts of use:

> Woodward's theory of causality licenses counterfactuals about what would happen if the cause were to vary in a very special way, a way that is not what we would normally be envisaging for either policy or technology. The variation is just the kind we need if we want to test a causal claim – to hunt for causes – but not the kind we expect to implement when we try to use them. (Cartwright 2007, p.48)

This is not surprising. Policy and technology are similar in the sense that these topics focus our philosophical attention to *use*.

As I argued above, in the philosophy of physics, the use-perspective is almost completely absent. Nevertheless, the examples I gave in 1.2.2 suggest that finding evidence for using causal claims is an endeavour worthy of attention in this area too. In this dissertation, I take this suggestion seriously, and take up the use-perspective to study physical causal claims. When we see technology on par with policy, physical causal knowledge that we use to design, use, maintain, repair, and improve artefacts becomes important.

1.5 Philosophy of technology on artefacts

Finally, I want to reflect a bit on a philosophical domain that shares a focus on artefacts with my dissertation, viz. philosophy of technology. Because of the focus that philosophy of technology lays on the study of artefacts, it might seem that my dissertation fits better in the philosophy of technology than in philosophy of science. Philosophers in the SPS have of course argued that this distinction is not as rigid as often thought. But regardless of that, while the focus on artefacts might be similar, the questions that I address differ significantly from those traditionally asked by philosophy of technology. Below, I present a brief review of important topics and points of discussion of the philosophy of technology. This provides a setup for the chapters and case studies to come. At the same time, it will help show how my dissertation differs from what is traditionally done in philosophy of technology. To make a study of artefacts useful, I first need to go into some important topics of what artefacts are and how to characterise them and the disciplines related to artefacts.

1.5.1 Artefacts are man-made

The first and most important point about artefacts in the context of this dissertation, is that they are man-made (Bucciarelli 2003, p.1). In this way, they are distinct from natural objects (Kroes 2012, p.3). Artefacts are the result of human interventions, while natural objects are not. However, a quick look around shows that it is not clear where to draw the line between "natural" and "artificial". Peter Kroes gives the example following example: "what kind of human work and how much of it" is needed to turn a stone into an axe (2012, p.14)? Kroes himself argues that such a distinction might be impossible. Fortunately, the specific boundaries between the two do not matter for my dissertation. What matters is the ways in which we use causal knowledge, which is something we do in relation to artefacts, but also in relation to natural objects and even in relation to possible hybrids. I will focus on the fact that artefacts are created: they are assembled in a specific way, with a specific goal. In philosophy of technology, this 'goal' is often related to the artefact's *function*.

1.5.2 Artefacts have a function & can malfunction

Philosophers of technology generally agree that having a function distinguishes artefacts from mere physical objects[14] (Kroes 2012, p.5). They also consider that the function is central to an artefact (Houkes and Vermaas 2010, p.2). Yet what this function is, or how to characterise it, is not agreed upon (Houkes and Vermaas 2010, p.2). Wybo Houkes and Pieter Vermaas distinguish three more or less traditional answers to the question who or what determines the technical function of an artefact: intentional theories, Robert Cummins' causal-role theory and evolutionary theories. On the first, human intentions "fix the functions of technical artefacts" (Houkes and Vermaas 2010, p.2). Functions are equated with intended effects. This is related to the 'goal' I mentioned above. One example of intentional theories is Kroes' account which sees functions as the "for-ness" of artefacts. A common problem with these theories is that it is unclear whose intentions count. If design-engineers intend for an artefact to be used in a specific way, but users

[14] Philosophers of technology do not really make explicit this contrast class of "physical objects" consists of. If they take it to include living organisms, that would construct a problem, since philosophers of biology are also rather concerned with functions and function ascriptions (see for instance (Sober 1993, section 3-7), (Godfrey-Smith 2013, p.2), (Garvey 2007, chapter 7)). However, works like (Krohs and Kroes 2009) give us good reasons to assume that philosophers of technology are aware of the debate in biology. So "physical object" seems to exclude living organisms.

have other intentions, which of these determine the function of the artefact? On the causal-role theory, functions roughly correspond to "the causal contribution the item makes to the system containing it" (Houkes and Vermaas 2010, p.2). Houkes and Vermaas argue that items can make many contributions to many different systems, and that those contributions do not necessarily contribute to their functions. On their view, this constitutes a problem for causal-role theories. Moreover, choosing a specific system or a specific contribution to focus on looks like a process related to human intentions, which brings us back to the problems associated with the intentional theories. The final set of theories Houkes and Vermaas discuss are evolutionary theories. Evolutionary theories disconnect function from human intentions, and see them as the result of "evolutionary forces like variation and selection" (Houkes and Vermaas 2010, p.2). Here, questions rise regarding which processes are relevant and what the role of directed and purposeful design can still be (Houkes and Vermaas 2010, pp.2-3).

As an alternative, Houkes and Vermaas formulate an account of functions that combines insights from all three traditional answers, while giving some priority to the intentionalist part (Houkes and Vermaas 2010, pp.2-3). It is aptly called the ICE-theory. It states that

> An agent a justifiably ascribes the physicochemical capacity to φ as a function to an item x, relative to a use plan up for x and relative to an account A, if:
>
> I. a believes that x has the capacity to φ;
> a believes that up leads to its goals due to, in part, x's capacity to φ;
>
> C. can on the basis of A justify these beliefs; and
>
> E .a communicated up and testified these beliefs to other agents, or a received up and testimony that the designer d has these beliefs.
>
> (van Eck 2016, p.4)

Houkes and Vermaas avoid problems associated with the aforementioned accounts by adopting a very liberal notion of design: both engineers, redesigners and creative users count as designers.

The specific notion of function that one adopts does not really matter for my dissertation. I mainly discuss this because it is an important topic in philosophy of technology, and because it is crucial to the notion of *design*, which is important for the chapters to come. I will go into the notion of design in the next subsection. First, let me discuss the importance of *malfunction.*

Malfunction refers to the situation where an artefact fails to perform its function (whatever it was). Several philosophers of technology have stressed that a definition of function should account for the possibility of malfunction (e.g. (Houkes and Vermaas 2010, p.3), (Floridi, Fresco, and Primiero 2015, pp.1200-1201), (van Eck 2016, p.9)). It

should also be distinguished from what Behrooz Parhami calls a defect, for example, which refers to something that is physically damaged but still functioning (1997, p.450). Parhami presents a multi-level model of reliability in which defect and malfunction are two stages, failure being the final one describing the malfunctioning of the entire system (1997, p.451). I am not concerned with the differences between failure and malfunction per se. I mainly want to draw attention to the idea that both are connected to the function of the artefact, more than to its physical state. Malfunction and failure are related of course, since the physical state of an artefact can give rise to malfunctioning, and often repairing the artefact will imply interventions onto the physical state. I will get into this in 1.5.4.

1.5.3 Artefacts are designed

Design is closely related to artefacts and to functions. On an intuitive level, design is the process with which we make sure that the artefact can perform its function. But, as was the case with functions, there is no consensus about the definition of design (Buchanan 2009). Kroes defines it as "processes in which functions are translated into structures" (Kroes 2012, p.27). He connects this to functional decomposition[15], an assessment where an artefact is represented in terms of the components' functions only, in a way that is relevant for the current purpose such as malfunction analysis (van Eck 2016, pp.12-13). Most philosophers connect design more with the making of plans, like Kees Dorst and Kees van Overveld:

> Design is a human activity in which we create plans for the creation of artifacts that aim to have value for a prospective user of the artifact, to assist the user in his/her effort to attain certain goals. (2009, p.4556)

According to Per Galle and Peter Kroes, design is the action of producing an idea for a new artefact, in such a way that it allows others to make artefacts according to that idea (Galle and Kroes 2014, p.216). This is also the way Houkes and Vermaas define designing, though in their view, plans are significantly more liberal. They view plans as "mental items that consist of considered actions", regardless whether these actions are actually carried out. In this way, they also accommodate amateur designing and common sense designing as proper design activities, though these are respectively not done by trained

[15]William Bechtel and Robert Richardson also discussed functional decomposition as a strategy for discovering mechanisms in the biomedical sciences (Bechtel and Richardson 1993, p.xxx).

and licensed professionals, or not based on specialised knowledge (Houkes and Vermaas 2010, p.27).

But regardless of the specific definition, the act of designing is still aimed at creating artefacts, objects with a certain function. In attempts to understand design practice, philosophers have studied whether the activity is best characterised as problem-solving (Buchanan (2009)), whether it is rational (Kroes (2009a), Franssen (2015) and Bucciarelli (2003)), how to reconcile different goals (de Vries (2009)),… I cannot go into these questions here. What I do want to stress is that several philosophers have drawn attention to the *synthetic nature* of design. Kroes, for example, links this to specific skills:

> […] engineers need to have synthetic design skills: when designing new technical artefacts, they must be able to combine elements (components or processes) in inventive, creative ways so that they can satisfy practical means-end or functional requirements. The designing of technical artifacts is considered to be primarily a synthetic rather than an analytic activity. […] For these purposes they also need to have synthetic skills; theories, experiments as well as experimental equipment are composed of different elements (like, for instance, laws, actions and physical components) and they result from researchers putting these elements together in specific ways to satisfy requirements, cognitive and otherwise. (2009b)[16]

According to Richard Buchanan, the characterisation of design as synthetic activity can be traced back to Herbert Simon and refers to the idea that in designing, engineers put things together to create a functioning whole:

> He designs by organizing known principles and devices into larger systems. (Buchanan 2009, p.425)

Although it is an intuitive notion, 'synthetic' can have many different meanings. Yet design is a synthetic activity in a very specific sense: it

> […] involves the synthesis of functional components that together realize the overall function of a technical artifact. (Kroes 2009b, p.406)

In the context of this dissertation, the most important questions are: How do we achieve this? How do we create a complex whole that (most of the time)[17] successfully performs the function we want it to? This brings us to the material realisation of artefacts.

[16] This does not mean that designing is solely a synthetic activity.

[17] Following (Floridi, Fresco, and Primiero 2015), if artefacts are incapable of ever performing their function, they are not failing but rather are badly designed.

1.5.4 Artefacts need to be realised materially

To create actual artefacts, our design plans need to be carried out physically: the artefacts need to be constructed and built. Kroes expresses this as the dual nature of artefacts: they cannot be captured by mere intentional conceptualisations since they need to be designed and physically made, but neither can they be conceptualised in mere physical terms, since artefacts have an intended function. (Kroes 2012, p.5) This making, or physical realisation of artefacts is not a straightforward thing:

> Matter has to be transformed so that the resulting physical construction has certain capacities or shows a particular kind of behavior. Often that is an arduous process which may involve many problems, setbacks, failures. (Kroes 2012, p.3)

Kroes is not the only one to stress the difficulty of actually creating artefacts. Hans Radder for example also reflects on the difficulty of building artefacts:

> First, we need the capability for, literally, putting together a technological system that has the potential of performing the required function. This means that we need to have available the materials, resources, skills and knowledge that are required for designing, constructing and using the technological system in the first place. (2008, p.54)

Because of this, Kroes argues that this putting together of components to realise an overall function of an artefact, is something that makes the designing of technical artefacts "a synthetic activity with distinctive features of its own" (Kroes 2009b, p.406).

1.5.5 Artefacts need to be stable and reproducible

The material realisation of artefacts needs to be such that the artefact can perform the intended function. Yet this is not all. Radder argues that, to be successful, the realisation needs to be reproducible, and have a certain degree of stability. It needs to be reproducible in the sense that "different systems of the same type should be able to exhibit the same function" (Radder 2008, p.53). So it should not be the case that we cannot build another instance of a design plan – we should be able to create more than one functioning copy[18]. Secondly, the artefact should be able to perform its function

[18] This mainly applied to artefacts in the area of mass production, and less to artefacts like custom made jewellery or art.

"across a variety of situations and during a substantial period of time" (Radder 2008, p.52). So a successful artefact should not break down every ten minutes or only work when it is exactly 25 degrees outside. The demands for reproducibility and stability are actually part of the reason why it is so daunting to successfully create a functioning artefact, and why design is so central. They are also part of the reason why my cases from failure analysis are so interesting. As will turn out, the knowledge from analysing failed artefacts can help improve future designs with regard to stability.

1.5.6 Maintenance

One important way to *keep* artefacts stable, is by *maintaining* them. This is silently assumed by philosophers of technology, but not often stressed or discussed. An important exception is Radder (2008) (which should not be a surprise, since he stresses stability). For my dissertation, it is important to sketch the image of an artefact as a complex arrangement of components organised in such a way that the resulting arrangement behaves in a particular way. This is not an easy task and requires constant attention and interventions throughout its lifespan to keep it functioning. After all,

> […] since the world may and often does change in substantial ways in the course of time, keeping a working technological system stable and reproducible also requires active and intentional human intervention. That is why technologies, if they are expected to keep functioning, cannot be left to themselves. (Radder 2008, p.54)

Nevertheless, maintaining artefacts to ensure their stability is an activity similar to building artefacts: it requires specific knowledge and skills. As I already sketched above and will argue throughout this dissertation, deciding which knowledge is applicable and getting to that knowledge is not a straightforward thing.

1.5.7 Artefacts are embedded in a physical and social environment

This demand also brings us to the next feature of artefacts, namely their relation to their environment. Artefacts need to function in a specific environment (viz. those things not included in the artefact that are nevertheless relevant for its functioning, see (Radder 2008, p.52)). This can be the physical environment (like the temperature mentioned above, but also surface, whether there is wind,…) but also the *social* environment. For my cases, the physical environment will be more important than the social, but nevertheless, social embedding is an important feature of artefacts and I want to spend some time on it.

As a first point, any material realisation is always part of a social context: there are always humans and human interactions, social structures etc. in which the artefact functions (Radder 2012, p.159).

This embedding can have significant influence. A clear way in which the social environment plays a role is in the way artefacts are *used*. Artefacts can be used in very different ways: a chair can be used for sitting, for standing on, even to hit people with. The latter is an unusual way of using the artefact and most often also illegal - two possible social limitations to artefact use (Houkes and Vermaas 2010, p.6). The artefact itself cannot enforce any behaviour or use: there needs to be a social system in place that enforces sanctions on incorrect or unwanted uses or interactions, like a justice system with laws (Kroes 2012, p.16). Correct or proper ways of using artefacts are often placed in void clauses to cover the makers when things go wrong[19] (Houkes and Vermaas 2010, p.6). This is related to the demand for stability: clearly, social environment also determines whether the artefact is considered stable (Radder 2008, p.55). If a common practice is to bang chairs against the floor e.g. for a celebration, then for chairs to be stable, the chair-design needs to take these actions into account.

A final important way in which society and artefacts are linked, is on a more ethical level. Producing artefacts is a costly business. Correspondingly, social agents with the most money (and connected to that, power) will determine which artefacts are made most frequently. They will also determine which artefacts are more researched than others. For the artefacts I will be talking about, this may not seem important: it's just machines. But look for example at pharmaceutical products. Big pharmaceutical companies can influence the research that is done in biomedical sciences in a significant way (Radder 2012, p.59). This is an important topic of reflection. Maybe less straightforward, this is also an important topic of reflection about machines. For example, if most research is done into nuclear energy because the industry funds this more than e.g. windmills or solar panels, this can seriously influence the development of technology and correspondingly, the development of our society.

Note that these are important topics of reflection, but that according to Radder, philosophers should not get sole say on the policy on this matter. According to him, democracy is central to philosophy of technology. I wholeheartedly agree. Because of the social embedding of technology and artefacts, people need to be correctly and

[19] Liability is becoming increasingly important because of automation. With self-driving cars, for example, there is the question of who should be held responsible in case of an accident: the technology manufacturer or the driver (Malinas and Bigelow 2016).

sufficiently informed about technologies, safety issues, ethical concerns, impact on society etc. To achieve this, philosophical reflection is crucial.

1.5.8 Engineering sciences: the science of (phenomena in) artefacts

Besides the focus on design and how central it is for the function of artefacts, philosophy of technology tends to focus on reverse engineering and malfunction analysis as engineering disciplines. The first is a practice where engineers analyse how an artefact achieves its general function or behaviour, often to redesign it (in a more efficient way) afterwards (van Eck 2016, p.22). In malfunction analysis, on the other hand, engineers attempt to explain why an artefact failed to perform a certain expected function (van Eck 2016, p.10). Mieke Boon and Tarja Knuuttila have argued that there is more to the engineering sciences. They define it as a practice that tries "through modelling to explain, predict or optimize the behaviour of devices, processes, or the properties of diverse materials, whether actual or possible" (Boon and Knuuttila 2009, p.687). As a matter of fact, they argue that design is part of engineering, and not of the engineering *sciences*. The distinction they press is best seen in terms of models: the models that engineering produces are significantly different from the models that the engineering sciences produce. A way to make it tangible, is to see that engineering mainly focuses on how artefacts perform their *function* – which was indeed a central concept for (1)design (let's make the artefact perform the function), for (2) reverse engineering (*how* does the artefact perform its function?) and for (3) malfunction (why doesn't the artefact perform its function?). The point that Boon and Knuuttila want to make, is that there is also modelling of phenomena that occur in artefacts with no or less focus on function. As a contrast, "the engineering sciences aim at both furthering the development of devices and materials meeting certain functions and optimizing them" (Boon and Knuuttila 2009, p.688). They do this by modelling the behaviour of artefacts and phenomena that occur in artefacts in terms of physical phenomena (Boon and Knuuttila 2009, p.687).

Since this dissertation is written from a Philosophy of Science in Practice-perspective, I do not feel that it is necessary to hammer on a sharp distinction between engineering and the engineering sciences. The two are intertwined both conceptually and even practically, in research and development labs like the one associated with the Phillips company, or the AT&T Bell lab (Crow and Bozeman 2005). The demand for a specific artefact often drives the questions that engineering science asks and the research conducted. And conversely, in redesign for example, knowledge from engineering sciences (about properties of materials for instance) can be used to improve the engineering design of a specific type of artefacts. What is important, I think, is to see that

what Boon and Knuuttila call the engineering sciences, is a complex practice. This helps to debunk the idea that artefacts (and applications in general) are rather straightforward consequences of the laws of physics :

> […] applying scientific laws for describing concrete phenomena usually requires idealizations, approximations, simplifications and ad-hoc extensions (e.g. Cartwright 1983, p.111). As a result, in technological applications predictions based on scientific theories are not at all straightforward since boundary conditions not accounted for in the theory may be involved. Scientific theories do not give rules on how to idealize, approximate, simplify and extend a scientific law in order to make it fit for concrete phenomena. Consequently, scientific approaches to understanding or predicting phenomena relevant to technological applications involves scientific modeling different from the way textbooks present the application of fundamental theories in the construction of mathematical models for concrete systems (e.g., by using Newtonian mechanics, thermodynamics, electricity and magnetism, or quantum mechanics). (Boon 2011, p.64)

I discuss this in chapters 3 and 5. But for now, it's important to see that the relation between artefacts and science is not at all straightforward.[20]

1.6 Overview of this dissertation

Now that I provided enough information, I can repeat my three main topics and explain how I will proceed to discuss them. In this dissertation, I study physical causal knowledge needed for and acquired in using, researching, building and maintaining artefacts[21], with regard to

(1) the meaning of causal claims about artefacts,

(2) the evidence needed for using this physical causal knowledge, and

(3) the relation of this physical causal knowledge to the laws of physics.

[20] Maarten Franssen (2015, p.239) made a similar point.

[21] Though I mainly limit the use-contexts to artefacts, my analysis also holds for natural contexts. I will therefore on occasion explain how it can be adapted.

As will become clear throughout the dissertation, all three are interrelated. Causal relations with different meanings demand different types of evidence, for example. Moreover, the relation of physical causal knowledge to the laws of physics will in some sense be present in every chapter. To study these three aspects of physical causal knowledge, I will combine insights from philosophy of science and philosophy of physics; from philosophy of technology and of the engineering sciences; from philosophy of the biomedical sciences; and from philosophy of the social sciences. This is due to several reasons.

For one, as my overview of the literature has shown, there are some gaps between philosophy of science and philosophy of technology and of the engineering sciences. Philosophy of physics does not look at applications of physical knowledge, or to the wider context of physics in general. When it comes to studying applications of knowledge, philosophy of biomedical sciences and of social sciences do a lot better. Nevertheless, physical knowledge is also applied to explain and intervene in the world, e.g. when building, using, maintaining and repairing artefacts. These artefacts form the main topic of philosophy of technology and of engineering (science). Yet discussions in this field do not really focus on more traditional philosophical questions, like how we gather knowledge to use in creating and maintaining artefacts or what evidence we need for this knowledge. I will attempt to bring these fields closer together.

Second, the focus on *use* is more prominent in philosophy of biomedical and of social sciences than in philosophy of physics. I will use and where necessary, adapt, the relevant contributions in the literature to shed light on my cases. The combination of these different philosophical fields hopefully may improve the status of all the involved fields, and contribute to a philosophical understanding of science in the broad sense.

I will not only use a great diversity of philosophical resources, but also a great diversity of case studies with increasing complexity. Specifically, I will look into technical repair manuals (see chapter 2), common artefacts (chapter 3) and failure analysis (chapters 4 and 5).

As I explained, each chapter will argue for one of the four specific claims that I formulated in the Introduction. In chapter 2, I will start with remedy instructions from technical repair manuals of bikes, cars and radios because they constitute a very day-to-day context where we find use-related physical causal knowledge. The manuals prescribe *interventions* that if performed, fix malfunctioning artefacts. In order for the interventions to be warranted, they need to be based on appropriate causal relations. I will investigate what properties are required of the causal relation to warrant the instructions. This chapter will provide an argument for my first specific claim (I).

In chapters 3 and 4, I will turn to evidence for physical causal claims. Chapter 3 will deal with causal claims in the context of *explaining* phenomena in artefacts. I will

investigate what evidence is needed, and will conclude that the laws of physics do not suffice as source of evidence. Mechanistic evidence, I will propose, is needed to supplement the evidence that laws provide. In this way, this chapter allows me to argue for my second specific claim, viz. (II).

In chapter 4, I will discuss a more complex practice, viz. failure analysis – a specialised part of the engineering sciences. I will show that engineers generalise causal knowledge from one failure, in order to improve design and maintenance practices of other and future artefacts. I will investigate how the evidence they require for these generalisations can be characterised. Based on the cases, it will be shown that the evidence mainly consists of mechanistic information about the artefact and the context. This chapter provides support for my claim (III).

In the fifth chapter, I will use a case from failure analysis to investigate the relation between applications and laws of physics. By analysing the regularities that failure analysts use, I will argue that the view that universal, necessary laws are more appropriate than other regularities to achieve our epistemic goals is untenable. In reality, a whole array of regularities is used to achieve goals, and which of those regularities is used depends on the context and the specific goals we have. This chapter argues for the final specific claim, viz. (IV).

As announced, I will also realise two more generic aims (viz. (A) and (B)) throughout this dissertation. In the conclusion of each of the following chapters, I will explain how the chapter contributes to establishing these more generic aims. As such the case for (A) and (B) will gradually develop. In the Conclusion of this dissertation, I will summarize my arguments for the four specific claims and reflect on how a combination of these arguments helps me reach my two more generic goals. The Conclusion also contains a reflection on the fruitfulness of a more practice and use based philosophy of physics.

Several of the following chapters are based on individual papers. However, I have made significant alterations and additions to improve the overall coherency and narrative of the dissertation. Some of the papers are co-authored. For the corresponding chapters, I have made related formal changes, such as writing from the “I” instead of “we”-perspective. Nevertheless, the material originated in cooperation, and I will therefore systematically specify co-authors where needed.

References

Allhoff, Fritz. 2015. Philosophies of the Sciences. In *Philosophies of the sciences: a guide*, edited by F. Allhoff: John Wiley & Sons.

Bechtel, William, and Adele Abrahamsen. 2005. Explanation: a mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (2):421-441.

Bechtel, William, and Robert C. Richardson. 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton: PUP.

Boon, Mieke. 2011. In Defense of Engineering Sciences: On the Epistemological Relations Between Science and Technology. *Techne* 15 (1):49-71.

Boon, Mieke, and Tarja Knuuttila. 2009. Models as epistemic tools in engineering sciences: a pragmatic approach. *International Journal of Software Engineering and Knowledge Engineering*:687-720.

Bucciarelli, Louis L. 2003. *Engineering philosophy*: Delft University Press.

Buchanan, Richard. 2009. Thinking about design: an historical perspective. In *Philosophy of Technology and Engineering Sciences*, edited by A. W. M. Meijers, D. M. Gabbay, P. Thagard and J. Woods. Amsterdam: North Holland.

Cartwright, Nancy. 1979. Causal laws and effective strategies. *Nous*:419-437.

———. 1983. *How the laws of physics lie*: Oxford University Press.

———. 1999. *The dappled world: essays on the perimeter of science*. Cambridge: Cambridge University Press.

———. 2007. *Hunting causes and using them: approaches in philosophy and economics*. Cambridge: Cambridge University Press.

Cartwright, Nancy, and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford, New York: Oxford University Press.

Causality in the Sciences. 2011. Edited by P. Illari, F. Russo and J. Williamson: Oxford University Press.

Craver, Carl F., and Marie I. Kaiser. 2013. Mechanisms and Laws: Clarifying the Debate. In *Mechanism and Causality in Biology and Economics*, edited by H.-K. Chao, S.-T. Chen and R. L. Millstein: Springer Netherlands.

Crow, Michael, and Barry Bozeman. 2005. *Limited by Design: R & D Laboratories in the U.S. National Innovation System*: Columbia University Press.

De Vries, Marc. 2009. Translating Customer Requirements into Technical Specifications. In *Philosophy of Technology and Engineering Sciences*, edited by A. W. M. Meijers, D. M. Gabbay, P. Thagard and J. Woods. Amsterdam: North Holland.

Dorst, Kees, and Kees van Overveld. 2009. Typologies of design practice. In *Philosophy of Technology and Engineering Sciences*, edited by A. W. M. Meijers, D. M. Gabbay, P. Thagard and J. Woods. Amsterdam: North Holland.

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

———. 2011. The causal-process-model theory of mechanisms. In *Causality in the Sciences*, edited by P. Illari, F. Russo and J. Williamson: Oxford University Press.

Eells, Ellery. 1991. *Probabilistic causality*. Vol. 1: Cambridge University Press.

Floridi, Luciano, Nir Fresco, and Giuseppe Primiero. 2015. On malfunctioning software. *Synthese* 192 (4):1199-1220.

Fodor, Jerry A. 2007. Special Sciences (or: The Disunity as a Working Hypothesis). In *Philosophy of science: an anthology*, edited by M. Lange.

Franssen, Maarten. 2015. Philosophy of Science and Philosophy of Technology: One or Two Philosophies of One or Two Objects? In *The Role of Technology in Science: Philosophical Perspectives*, edited by S. O. Hansson. Dordrecht: Springer.

Freeman, Christopher, and Luc Soete. 2009. Developing science, technology and innovation indicators: What we can learn from the past. *Research Policy* 38 (4):583-589.

Frisch, Mathias. 2014. *Causal reasoning in physics*: Cambridge University Press.

Galle, Per, and Peter Kroes. 2014. Science and design: Identical twins? *Design Studies* 35 (3):201-231.

Garvey, Brian. 2007. *Philosophy of biology*: Stocksfield : Acumen.

Giere, Ronald. 1997. *Understanding Scientific Reasoning*. Forthworth: Harcourt Brace College Publishers.

Gillies, Donald. 2011. The Russo-Williamson thesis and the question of whether smoking causes heart disease. In *Causality in the Sciences*, edited by P. Illari, F. Russo and J. Williamson: Oxford University Press.

Glennan, Stuart. 2011. Singular and general causal relations: A mechanist perspective. In *Causality in the Sciences*, edited by P. Illari, F. Russo and J. Williamson: Oxford University Press.

Glymour, Clark N. 2001. *The mind's arrows: Bayes nets and graphical causal models in psychology*: MIT press.

Godfrey-Smith, Peter. 2009. *Theory and reality: An introduction to the philosophy of science*: University of Chicago Press.

———. 2013. *Philosophy of Biology*: Princeton University Press 2013.

Hausman, Daniel M. 1998. *Causal Asymmetries*: Cambridge University Press.

Hitchcock, Christopher Read. 2004a. Are there laws in the social sciences? In *Contemporary Debates in the Philosophy of Science*, edited by C. Hitchcock: Blackwell Publishing Ltd.

———. 2004b. What is the Philosophy of Science? In *Contemporary Debates in the Philosophy of Science*, edited by C. Hitchcock: Blackwell Publishing Ltd.

Houkes, Wybo, and Pieter E. Vermaas. 2010. *Technical functions: On the use and design of artefacts*. Vol. 1: Springer Science & Business Media.

Illari, Phyllis, and Federica Russo. 2014. *Causality: Philosophical Theory meets Scientific Practice*. 1 edition ed: Oxford University Press.

Illari, Phyllis, and Jon Williamson. 2011. Mechanisms are real and local. In *Causality in the Sciences*, edited by P. Illari, F. Russo and J. Williamson: Oxford University Press.

Kincaid, Harold. 2004. There are Laws of the Social Sciences. In *Contemporary Debates in the Philosophy of Science*, edited by C. Hitchcock: Blackwell Publishing Ltd.

———. 2011. Causal modelling, mechanism, and probability in epidemiology. In *Causality in the Sciences*, edited by P. Illari, F. Russo and J. Williamson: Oxford University Press.

Kroes, Peter. 2009a. Foundational Issues of Engineering Design. In *Philosophy of Technology and Engineering Sciences*, edited by A. W. M. Meijers, D. M. Gabbay, P. Thagard and J. Woods. Amsterdam: North Holland.

———. 2009b. Introduction to part III. In *Philosophy of Technology and Engineering Sciences*, edited by A. W. M. Meijers, D. M. Gabbay, P. Thagard and J. Woods. Amsterdam: North Holland.

———. 2012. *Technical Artefacts: Creations of Mind and Matter*. Vol. 6, *Philosophy of Engineering and Technology*. Dordrecht: Springer Netherlands.

Krohs, Ulrich, and Peter Kroes. 2009. *Functions in biological and artificial worlds : comparative philosophical perspectives*: Cambridge (Mass.) : MIT press.
Leuridan, Bert, and Erik Weber. 2011. The IARC and Mechanistic Evidence. In *Causality in the Sciences*, edited by P. Illari, F. Russo and J. Williamson: Oxford University Press.
Lewis, David K. 1973. *Counterfactuals*. Cambridge: Harvard University Press.
Mach, Ernst. 1901. *Die Mechanik in ihrer Entwickelung historisch-kritisch dargestellt*. Leipzig: Brockhaus.
Mackie, John L. 1980. *The Cement of the Universe*. 2nd ed: Oxford University Press.
Malinas, Gary, and John Bigelow. 2016. Simpson's Paradox. In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta: Metaphysics Research Lab, Stanford University.
Norton, John D. 2006. Do the Causal Principles of Modern Physics Contradict Causal Anti-Fundamentalism? In *Thinking about causes: from greek philosophy to modern physics*, edited by P. K. Machamer and G. Wolters: University of Pittsburgh Pre.
———. 2007. Causation as Folk Science. In *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, edited by H. Price and R. Corry: Oxford University Press.
———. 2009. Is There an Independent Principle of Causality in Physics? *The British Journal for the Philosophy of Science* 60 (3):475-486.
Norton, John D. 2008. The Dome: An Unexpectedly Simple Failure of Determinism. *Philosophy of Science* 75 (5):786-798.
The Oxford handbook of causation. 2009. Edited by H. Beebee, C. Hitchcock and P. Menzies: Oxford University Press.
Parhami, Behrooz. 1997. Defect, Fault, Error, . . . or Failure? *IEEE Transactions on Reliability* 4 (46).
Pearl, Judea. 2009. *Causality*: Cambridge university press.
Price, Huw, and Brad Weslake. 2009. The Time-Asymmetry of Causation. In *The Oxford Handbook of Causation*, edited by H. Beebee, P. Menzies and C. Hitchcock: Oxford University Press.
Psillos, Stathis. 2002. *Causation and explanation*. Vol. 8: McGill-Queen's Press-MQUP.
———. 2007. *Philosophy of science AZ*: Edinburgh University Press.
Radder, Hans. 2008. Critical Philosophy of Technology: The Basic Issues. *Social Epistemology* 22 (1):51-70.
———. 2012. *The Material Realization of Science: From Habermas to Experimentation and Referential Realism*. 1 ed, *Boston Studies in the Philosophy of Science 294*: Springer Netherlands.
Roberts, John T. 2004. There are no Laws of the Social Sciences. In *Contemporary Debates in the Philosophy of Science*, edited by C. Hitchcock: Blackwell Publishing Ltd.
Russell, Bertrand. 1912. On the notion of cause. *Proceedings of the Aristotelian society* 13:1-26.
Salmon, Wesley C. 1998. *Causality and Explanation*: Oxford University Press.
Schaffer, Jonathan. 2016. The Metaphysics of Causation. In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta: Metaphysics Research Lab, Stanford University.
Schurz, Gerhard. 2013. *Philosophy of science: A unified approach*: Routledge.
Shapin, Steven. 2016. Invisible Science. *The Hedgehog Review* 18 (3).
Sober, Elliott. 1984. Two concepts of cause, 1984.
———. 1986. Causal factors, causal inference, causal explanation. *Proceedings of the Aristotelian Society, Supplementary Volumes* 60:97-113.

———. 1993. *Philosophy of biology*: Oxford : Oxford university press.
Suppes, Patrick. 1970. *A probabilistic theory of causality*: North-Holland Publishing Company Amsterdam.
van Eck, Dingmar. 2016. *The Philosophy of Science and Engineering Design*: Springer.
Van Fraassen, Bas C. 2008. *Scientific Representation: Paradoxes of Perspective*. Reprint edition ed: Oxford University Press.
Woodward, James. 2003. *Making Things Happen*: Oxford University Press.

Chapter 2 Causation and Technical Problem Solving[1]

In this chapter I will investigate a very day-to-day context where we find use-related physical causal knowledge: in repair manuals of artefacts such as bikes, cars and radios. The manuals prescribe *interventions* that if performed, fix malfunctioning artefacts. In order for the interventions to be warranted, they need to be based on appropriate causal relations. Because of our familiarity with the artefacts in question, it might seem that fixing them, or producing the appropriate causal knowledge to fix them, is not that hard. Recall from 1.2.3 that Norton also assumed that using physical causal knowledge is straightforward and that it does not raise interesting philosophical questions. Here, I will show that this is mistaken and that producing practically useful causal knowledge, even for such seemingly simple artefacts as bikes, cars and radios, is rather complicated.

I will discuss three complicating factors: (1) causal knowledge is context- (and artefact-) dependent, (2) depending on the demands we place on our intervention we require different causal relations to hold, and (3) the artefacts on which we intervene are embedded in physical and social environments. My discussion of the first complicating factor will result in a tool to specify the domain of causal knowledge or claims. To investigate the second factor, I will first present some criteria for successful manuals. I take these criteria to be implicit in the practice of writing manuals, and I aim to make them explicit. These criteria will determine the meaning of the causal knowledge that is needed to underpin the instructions. I will then combine insights from Ronald Giere, Ellery Eells and John Mackie to capture the resulting meaning of the causal knowledge.

[1] This chapter is based on a paper co-authored with Erik Weber. I am grateful to four anonymous referees for helpful comments and reflections.

The third complicating factor is inspired by the debates in philosophy of technology that I described in 1.5.3 and 1.5.7.

Based on these three factors, I will argue that using and producing physical causal knowledge, even in such seemingly obvious cases, requires information about the context (both in a limited and in a broader sense), about the meaning of the causal knowledge and about how this relates to the way we want to use it. This chapter thus constitutes an argument for my first specific claim, viz. (I). At the same time, this chapter shows that focusing on useful physical causal knowledge does raise important philosophical questions, e.g; about the meaning of the physical causal claims to be made and how this meaning relates to how these claims are used. In this way, this chapter forms the first step in my exposition to reach my two generic aims (A) and (B). In the following chapters, I gradually introduce more complex cases and with that, more and more complications and interesting philosophical questions will arise.

Introduction

Three complicating factors

In this chapter I will focus on one of the ways in which we use physical knowledge that I defined in 1.2.2, namely for *interventions*. Woodward (2003) gave a technical definition of an "intervention" that is the standard in the current literature. I use this in a more colloquial meaning. I will show that determining whether certain causal claims about phenomena are adequate to form a base for intervening on these phenomena, is not an easy matter. To study the causal knowledge required for interventions, I will look at examples from technical problem solving manuals, aimed at allowing non-experts to repair their cars, bikes or radios. As I will show, these manuals prescribe certain actions or interventions, which are supposed to solve the malfunctioning. By studying the physical causal knowledge needed to warrant the prescribed interventions, several complicating factors become evident. I will discuss three such factors that are present even when we discuss rather reliable causal knowledge[2] about rather easy artefacts.

[2] The causal knowledge is reliable because it is given to us in manuals written by experts, so we do not need to support it ourselves (see chapter 3).

One, the causal relation, among which the quantitative relata and even the causal direction, can differ depending on the context[3]. While I agree with Norton that heat causes thermal expansion and not vice versa[4], for many other physical regularities, the causal direction can switch depending on the circumstances or the artefact. Causal knowledge can only guide our interventions if the domain of the causal knowledge corresponds to the situation you are intervening in. Correspondingly, evaluating causal claims in the context of interventions is a context-dependent activity. Of the three complicating factors, this one is the most discussed in the literature.

Two, it is not clear what it *means* when we say that x causes y, and the meaning can differ depending on the circumstances. As I explained in 1.3.1, many definitions of "cause" have been presented throughout the literature. Yet, the specific properties that a causal relation has, determine which actions one can warrantedly perform based on it[5]. Correspondingly, when we want to apply or produce appropriate causal knowledge, it's important to know what causal relations mean and which properties of the relation we need in order to warrantedly base our interventions on. To understand this general point, consider the difference between probabilistic causal relations and deterministic ones. If we want to be certain that our intervention always produces the wanted effect[6], we need to base it on a deterministic causal relation. Probabilistic relations will not do. So to intervene in a warranted way, we need to know the meaning of the causal relation on which we base our intervention.

Three, the (parts of the) artefacts we intervene on are embedded in broader physical and social contexts that also influence the effects of these interventions. This complication is related to the first one, but deserves separate attention. While the first complication focuses on the direct physical environment, this complication addresses the physical environments in a more broad way and emphasises social aspects. This complication is least discussed in the literature, and it is situated on a different conceptual level than the other complicating factors, so I will leave this for last.

These three complications form the starting point of the argument that I develop throughout this dissertation, viz. that the use of physical causal knowledge is a complicated and intriguing topic, worthy of philosophical attention.

---

[3] I will pay more attention to how the quantitative relata change in chapter 3.

[4] This is not that surprising, since thermal expansion is defined as expansion due to thermal energy (Considine and Kulik 2008, p.221).

[5] And, correspondingly, the evidence you need to support the causal relation. See also chapter 3 and 4.

[6] I will get back to this in 2.5.

In the next section, I will present the examples that I will use as cases throughout this chapter. In the rest of this introduction, I will present an overview of what is to come.

Examples

Everyone is in some sense familiar with what I call *technical problem solving instructions* (TPSI for short). Though you might not be familiar with the language, you probably are familiar with the practice of following instructions from a manual in order to fix an artefact. Here are some examples that are taken from car - and bicycle repair manuals:

(E_1) Excessive fuel consumption: the air filter element is dirty or clogged. Remove the air filter, clean with compressed air and refit. (Mead and Legg 1997, ref.15, 1.13)

($E_{1'}$) Excessive fuel consumption: the tyres are underinflated. Check and adjust pressures. (Mead and Legg 1997, ref.15, 0.14)

(E_2) The major cause of slow engine cranking or a "no-start" condition is battery terminals which are loose, dirty, or corroded. […] disconnect the battery and clean the terminals of both the battery and the cables. (Chilton 1986, p.16)

(E_3) Difficulty engaging gears: worn or damaged gear linkage or gear cable. Replace the cable. (Mead and Legg 1997, ref.15)

(E_4) Indication: Engine sputters, may fail to start. Condition: water in the fuel. Remedy: […] For a layer of water, the tank must be drained, and the fuel lines blown out with compressed air. (Chilton 1986, p.288)

(E_5) Starter motor turns engine slowly: partially discharged battery. Recharge. (Strasman 1988, p.23)

(E_6) When pedaling forward, the cassette spins, but there is no drive to the bike: the freehub body is worn. Replace the freehub body. (Sidwells and Ballantine 2004, p.37)

(E_7) The brakes are hard to apply, and/or sluggish to release: grit and dirt is inside the cable outers or the lubrication on the inner cables has dried. Strip down the cables, flush the outers, and clean the inner cables with degreaser, lubricate both, and reassemble. (Sidwells and Ballantine 2004, p.37)

To ensure that the analysis is not limited to ways of conveyance, I also include an example about radios:

(E_8) If the […] cone is badly torn or warped […] then the sound produced by the speaker will be distorted. Replace the loudspeaker with a new speaker, or recone the old one if a new model is not easily available. (Carr 1990, p.203)

The example about radios is in some sense outdated, since most contemporary radios are constructed in such a way that you cannot fix them yourself. However, they are paradigmatic in the history of (repairing) household electronics. At some point almost every household owned one, and when they malfunctioned, they needed to be repaired. Current versions of such an artefact are for instance printers or washing machines.

Based on these examples, I argue that a paradigm form of a TPSI contains the following three elements:

- a problem (e.g. excessive fuel consumption)

- a diagnosis (e.g. dirty air filter element)

- a remedy (e.g. remove, clean and refit the air filter).

Note that in some cases (e.g. E_7), there is more than one possible diagnosis.

In this chapter I will mainly focus on the third element of TPSIs: I will investigate the causal underpinnings of remedy claims. These claims prescribe an *intervention* on the malfunctioning artefact. On occasion, a manual also prescribes *diagnostic actions*: it instructs you to perform a certain action and observe the result. It then tells you what the result means in terms of diagnosing the specific malfunction. Examples are "test the state of charge of the battery using an individual cell tester or hydrometer" (Chilton 1986, p.279) and "when you apply the front brake and push the bike forwards, the headset moves forwards relative to the head tube" (Sidwells and Ballantine 2004, p.36). However, most of the time, these diagnostic actions are separated from the remedy instructions: they are in different sections, or in a different format etc. In this chapter, I will not focus on diagnostic actions, since they require a different framework of analysis than remedy instructions. In analysing the remedy instructions, I will take the relevant diagnostic actions to be performed and the outcome such that the diagnosis in the TPSI is established.

Structure of the chapter

In this chapter I will use the examples above to discuss three complicating factors for Norton's claim that ascribing causal relations to physical phenomena is straightforward. I will first spend time on showing that the causal relation that holds depends on the specific situations you are discussing. Hence, the validity of the corresponding causal claims differs too. Causal knowledge can only guide our interventions if the domain of the causal knowledge corresponds to the situation you are intervening in. In section 2.1 I will discuss this first complicating factor, and propose a way of dealing with it in our causal analysis.

I will then move on to the second complicating factor. This makes up the main part of this chapter (viz. sections 2.2 till 2.6). In section 2.2, I will provide some methodological reflection and define three criteria for the prescribed interventions to be successful: an efficiency requirement, a no harm requirement and an ideal of maximal assistance. I take these criteria to be implicit in the practice of compiling manuals. The subsequent sections can be seen as a stepwise analysis of how the abovementioned criteria determine the properties of the causal knowledge needed to underpin the proposed interventions. This will reinforce my suggestion that what it means for x to cause y is not a straightforward (let alone a single thing) thing. In section 2.3, I will investigate what kind of causal knowledge is necessary to satisfy the efficiency requirement. This will turn out to be the minimal strength of causal knowledge needed to warrant the remedy instructions and corresponding interventions: if the kind of knowledge defined in this analysis is absent, the efficiency requirement is violated. In section 2.4, I will then analyse how the no harm criterion (combined with the efficiency requirement) may demand knowledge of the presence of a stronger type of causal relation. In section 2.5, I will analogously examine how incorporating the maximal assistance ideal affects the kind of causal relations that are desirable. In section 2.6 I present a summary of my analysis. Because they will be useful for other chapters of this dissertation as well, section 2.6 also contains an overview of my definitions.

Finally, in section 2.7 I will pay some attention to the third complication: the physical and social embeddedness of interventions and artefacts can also influence the outcome of our interventions. When we intervene in in artefacts outside shielded, laboratory settings, these environments matter. However, they are often neglected or presumed to be of a certain kind. As I will argue, when intervening in real artefacts, they cannot be ignored: in order to find appropriate causal knowledge to base our actions on, we also need information about these environments.

This chapter will show that even for relatively easy artefacts, in relatively every day contexts, complications arise when we want to put causal knowledge to use. As I

mentioned, in the next chapters, I will show that as the situations become more complex or less familiar and the causal knowledge becomes less established, more and more complications arise. In this way, all the following chapters expand the argument that I will start here, viz. that using physical causal knowledge is complex and requires philosophical attention.

2.1 The First Complicating Factor

This section deals with the first complicating factor: causal relations are context-dependent and therefore the validity of corresponding ascriptions or causal claims is as well. I will explain how this context-dependency creates a difficulty for Norton's presumption that ascribing causal relations is straightforward. At the same time, I will argue why it is important to incorporate this context-dependency in an analysis of causal knowledge. I will therefore develop a tool to specify the domain of physical causal claims – an important aspect to analyse their meaning and validity.

2.1.1 Causal relations are context-dependent

Recall the problem of external validity I mentioned in 1.4 in relation to the biomedical and social sciences: that a causal claim holds in one context does not mean it will hold in another. Philosophers have therefore argued that both the validity and meaning of causal claims in the biomedical sciences are intrinsically linked with the *population* that the causal claim is about. I illustrate this with an example from Daniel Steel (2007, p.82). The following causal claims are true:

> Aflatoxin B1 causes liver cancer in rats.
> Aflatoxin B1 causes liver cancer in humans.

However, the following claim is false:

> Aflatoxin B1 causes liver cancer in mice.

Since the population we talk about makes a difference (not only in this case, but in many other cases as well), it is important that we always explicitly mention the intended population when making a causal claim. If we only say

> Aflatoxin B1 causes liver cancer

it is not clear which population we are talking about. And depending on which population we intend to talk about, the truth value of the claim differs.[7] This is especially important when we are looking for causal knowledge to base interventions on. Giving Aflatoxin B1 to mice will not result in cancer, giving it to humans will.

It is important to make the same kind of specification regarding physical causal claims. Without specifying what our causal claim is about, we cannot understand or evaluate it. Consider the following examples of general physical causal claims:

(C_3) Increasing the temperature of a gas causes an increase in volume occupied by the gas.

(C_4) Increasing the temperature of a gas causes an increase in volume occupied by the gas in rigid, closed containers.

(C_5) Increasing the temperature of a gas causes an increase in volume occupied by the gas in non-rigid or open containers.

C_4 is incorrect. Gas in a rigid, closed container cannot expand. So in rigid, closed containers, an increase in temperature of the gas will not lead to an increase in volume occupied by the gas. C_5, on the other hand, is a correct causal claim. We therefore cannot evaluate C_3 as such: it is underspecified. And if we are looking for causal knowledge to increase the volume of a gas, C_3 will not suffice. Note that this example already shows that Norton's treatment of causal ascriptions can be specified – and needs to be specified if we want to understand how causal knowledge can guide interventions. An important way in which we can specify is regarding *contexts*: determining what the causal relation is depends on the context. Moreover, temperature can also change due to volume changes when compressing a gas for example (see 3.1.4). So without knowing the context, we cannot determine what the causal relation is. Correspondingly, we cannot judge whether a causal claim provides an adequate base for interventions, without knowing which situations the claim refers to. Note that in the example above, not the entire context is specified. The colour of the container, its shape and the specific material it's made of, are for example not mentioned. This is because only the relevant contextual factors need to be taken into account. Determining what the relevant factors are, is not easy. I will pay more attention to this in chapter 3 (see 3.1.2 and 3.1.4).

[7] Peter Menzies (2007) and Christopher Hitchcock (1996) have also connected causation to contexts. I discuss them in 2.3.3.

So if we want to evaluate causal claims in a systematic way, we need a concept that can play the same role as 'population' does in causal analysis of the biomedical sciences, viz. specifying what the causal claim is about.

2.1.2 Physical setups

The concept I propose for this purpose is called 'physical setup'. It is defined as follows:

> A physical setup is a whole comprising at least two physical objects, located in space and time, with each having at least one variable feature.[8]

Consider an example of a physical setup based on example E_4 above:

(S_1) (1) The engine (with as variable feature whether it sputters or not) and (2) the fuel-system (with as variable feature whether it contains water or not), with (1) and (2) organised such that fuel from the fuel tank is pumped into the engine via the fuel system and located in time.

Physical causal claims can be seen as referring to physical setups in the following way: general (type level) physical causal claims are about collections (types) of physical setups; token physical causal claims (which I do not discuss in this chapter) are about individual setups.

By means of the concept of physical setup I can formulate the causal information underlying the remedy claim from E_4 as follows:

(C_6) For all physical setups of the type S_1: the value of the variable W (whether it contains water) of the fuel-system immediately before t influences/has an effect on the value of the variable S (whether it sputters) of the engine immediately after t.

Time t refers here to the moment at which the user of the car attempts to start it (i.e. turns the ignition key). The concept of physical setup now allows me to delineate what the causal claim is about. It is a tool to get a grip on the scope of causal claims. This concept therefore allows me to deal with the first complication. Whether a causal claim provides an adequate base for performing interventions depends on whether the domain of the causal claim corresponds with the situation you want to intervene in.

[8] "Physical object" is to be interpreted in a pragmatic way. It bears no metaphysical implications.

I now turn to explaining and handling the second complication. To that end, I will analyse the meaning of the remedy claim from E_4 and similar causal claims more thoroughly: how can "influences/has an effect on" be further characterised in light of the interventions that the claim intends to warrant? In Sections 2.3 – 2.5 I will gradually develop my answer and illustrate it with examples. To understand the upcoming analysis properly, I will first explain the general methodology of these sections in section 2.2.

2.2 The Second Complicating Factor – Methodology

In this section I will explain my methodology for the upcoming sections. They engage with the second complicating factor I mentioned in the introduction: the required properties of the causal relation we need to warrant interventions depend on what we want to achieve with the intervention. To study this, I will focus on the remedy claims of the TPSIs from the introduction. As should be clear from the examples, remedy claims are themselves not causal claims: grammatically they are imperative clauses (see the examples: "remove', "recharge", "adjust", ….). However, they have to be based on knowledge of certain causal relations that hold in the world, in order for the prescribed instructions to be warranted. Hence, I will not analyse causation in remedy claims (because there is no such thing) but causal knowledge underlying remedy claims.

2.2.1 Three criteria of success

The question that I will be answering in the following sections is "Which causal knowledge do we need to warrant remedy interventions like the ones prescribed in TPSIs?". The situation is often reversed: we want to intervene in a certain way, to reach a specific goal and we thus look for appropriate causal knowledge that can facilitate the intervention. In the case of the TPSIs, the interventions are prescribed, and I will analyse which causal knowledge is required to make this prescribing warranted. In general, repair manuals have a certain authority, since they are put together by experts. This gives us reason to trust that the knowledge needed to warrant the intervention is available. In chapter 3, I will discuss the complexities that arise when we do not have such a clear and trustworthy information source. So my point in this section is not to evaluate whether the writers of the manuals have the required knowledge. Rather, I will reflect on the demands that interventions and their goals put on the required causal

knowledge. To do so, we first need to have some idea of what the goal of these remedy interventions is. This might look straightforward: the goal is to fix the broken artefact. That is true, in a sense, but more needs to be said on the matter.

To that end, I will first present three criteria regarding the prescribed interventions that, in my view, a repair manual must satisfy to be successful. These criteria are implicit in practice, and by making them explicit, I can reflect on their properties and consequences for the manuals. As I will argue, each criterion requires the causal relation to have different properties to underpin the instruction. These criteria should not be regarded as strict rules: there can be exceptions.

The first criterion I propose is that repair manuals should avoid prescribing useless actions, i.e. actions whose result does not contribute to solving the problem. In order to clarify this criterion, it is useful to distinguish between the immediate result of an action and possible further consequences. Suppose I experience the temperature in my office as too high. I decide to open the window because I want a cooler room (that is my aim). The open state of the window is the immediate result of my action. Whether this immediate result leads in its turn to the desired state (a cooler room) after some period of time depends on an additional factor, viz. the outside temperature. If the room cools down, then I call this a consequence of my action. The useful distinction between results and consequences of actions has a history in the philosophy of action as well as in the literature on causation.[9] With this terminology in place I now formulate the *general efficiency requirement*:

> Including a remedy "Do X" is suitable only if there is a causal relation between R (the immediate result of the doing X) and the problem stated.

This general requirement has a specific instantiation for each TPSI. Let me give some examples. For E_5 and $E_{1'}$, the specific efficiency requirements are, respectively:

> Including the remedy "recharge the battery" is suitable only if there is a causal relation between battery charge and the speed at which starter motors can make the main engine run.
>
> Including the remedy "inflate the tyres" is suitable only if there is a causal relation between tyre pressure and fuel consumption.

[9] The distinction between results and consequences as I use it here stems from the work of Georg Henrik von Wright (1971, pp.66-67).

If a remedy claim does not satisfy the general efficiency requirement, it prescribes a useless action. Surely, making users waste time in performing useless acts is not a good idea.

The second criterion I propose is a *no harm requirement*: executing the instructions in repair manuals should not make the problem worse. Or, phrased in the terminology that I used above: the consequences of the actions should not be harmful. Exceptions are remedy instructions that function as a 'last measure': in cases where the artefact is in really bad shape, the manual might prescribe an action that if successful, fixes the artefact, but if unsuccessful, breaks beyond repair anyway, at least for the TPSI user – I will get back to this in 2.7. In such a case, a manual might prescribe an action that, either solves the problem, or further damages the artefact, within certain boundaries. However, since the artefact was not fixable by using the manual anyway, this action is not necessarily problematic. But on average, prescribed actions should not create new problems.

Finally, I argue that repair manuals should help the users as much as possible – within the range of actions they are capable of performing easily – in solving the problem(s) that the users experience. I call this the *maximal assistance ideal.*

These three criteria are a way of making explicit what we expect of the prescribed interventions. In that sense, they help decide which causal relations need to be in place (or by experts believed to be in place) to warrant the interventions.

2.2.2 A note on methodology

To analyse how these criteria influence the required properties of the causal relation that is needed to underpin the remedy claims, I will work with a series of definitions that capture different types of causation. The definitions are based on various existing theories of causation: the comparative model of Ronald Giere (1997), the context unanimity theory of Ellery Eells (1991) and the INUS theory of John Mackie (1980). I will extract interesting ideas from these theories and incorporate them into definitions that have a certain standard format. From Ronald Giere, I will extract the ideas of *positive causal factor* and *negative causal factor* to capture a very basic and very weak notion of causation. This will help me formulate the most basic requirement that I made explicit, viz. the efficiency requirement (section 2.3). This notion of causation is too weak to capture the no harm requirement. In order to capture this requirement, I will add the idea of *context unanimity* by Ellery Eells to the definitions from 2.3. The resulting definition specifies a causal relation that cannot be reversed in subcontexts – a stronger notion of causation. The ideal of maximal assistance also requires a stronger notion of

causation. To capture this, I will add the idea of *INUS condition* to the definitions of 2.3. This idea is inspired by John Mackie.

Combining ideas from three different philosophical theories on causation is not an easy task. All of these philosophers intended their theory to capture causation, in a monist way. As such, their complete accounts are not directly applicable to my cases. But they are suited for my pragmatic project that combines the ideas without the rest of their theoretical framework. To combine these ideas, I will reformulate them as definitions in a standard format. A template form of definitions in my standard format is the following:

> C (as opposed to C*) is a [causal notion] for E in the collection of setups U if and only if [criterion].

The causal notion is for example *positive causal factor*, like D_1 in 2.3, and the criterion captures the necessary and sufficient conditions that correspond to that causal notion. There is also mention of a contrast cause and of the domain (viz. the collection of setups, see 2.1.2). By formulating the ideas in this way, the relations between my definitions (how and why one is stronger than the other) become clear. My standard format also has several other advantages:

(I) The relevant type of physical setup is explicitly mentioned (so the tools are in place to avoid underspecification in the aforementioned sense).

(II) A definition in the standard format is "purely consequential", i.e. it specifies what follows from causal beliefs without making any assumptions about how causal claims are confirmed. In this way the definitions are compatible with the different ways in which evidence for causal claims can be gathered.[10]

(III) A definition in the standard format makes clear why causal knowledge is – in principle – practically useful (in the case of TPSIs: for solving problems, evaluating the proposed procedures, …). This will become clear in section 2.3.3.

(IV) My standard format takes into account the fact that the truth of causal claims often depends on the alternative cause we have in mind. I illustrate this with an example from Peter Menzies (2007, pp.204-206).[11] Consider three options for administering a drug to a patient: no dose, a moderate 100 mg dose, or a strong 200 mg dose.

[10] See chapters 3 and 4 for more information about evidence for causal claims.

[11] The example is originally from Christopher Hitchcock (1996), but he uses it to argue that causation is a ternary relation.

Suppose we give the patient a moderate dose and he recovers. It is possible that both of the following claims are true:

> Taking the moderate dose (as opposed to no dose) was a cause of the patient's recovery.
>
> Taking the moderate dose (as opposed to the strong dose) was not a cause of the patient's recovery.

This example is based on the idea that causes make a difference for their effects. Yet whether a cause is seen as making a *difference*, depends on the alternative that is considered. In the example, taking the moderate dose and the strong dose both cause the patient's recovery. So in the second case, taking the moderate dose did not make a difference for the patient's recovery. As such, it seems that taking the moderate dose was not a cause of the recovery. If we leave out the specification "as opposed to", we get two claims that seem to contradict each other but maybe are compatible, since they refer to different considered alternatives. To rule out this kind of confusion, definitions should explicitly mention the alternative cause. The same holds for causal claims regarding TPSIs. Suppose that we experience difficulty braking when driving our bike. When we are referring to the type of physical setup S_2:

(S_2) (1) The brakes (with as variable feature whether they are sluggish to release) and (2) the cable (with as variable feature whether it is gritty, clean or lubricated), in all cable operated bicycle brake systems.[12].

both causal claims can be true:

> The cable being clean (as opposed to lubricated) is a cause of the braking difficulties.
>
> The cable being clean (as opposed to gritty) is not a cause of the braking difficulties.

Again, if we do not specify what state we consider as the alternative, we get two claims that are contradictory.[13]

[12] Cable operated brake systems encompass all brake systems where pulling the cable engages the brakes, viz. makes the bike brake. This includes e.g. rim brakes and roller brakes.

2.3 The Efficiency Requirement

In this section I will investigate what kind of causal knowledge is necessary to satisfy the efficiency requirement. In section 2.3.1, I will first introduce Ronald Giere's comparative model for causation. This inspired my first two definitions. In section 2.3.2, I will present these definitions: positive causal factors (PCF) and negative causal factors (NCF). I will also reflect on how my definitions relate to Giere's comparative model of causation. In Section 2.3.3 I will show that my definitions have the properties specified in 2.2.2. I will also reflect on why I chose Giere's model as a basis for my definitions. Finally, in 2.3.4 I will argue that, in order to satisfy the efficiency requirement, remedy claims must be based on either positive causal factors or negative causal factors as defined in 2.3.2.

2.3.1 My use of Giere's comparative model

My first two definitions are based on the comparative model which Giere developed mainly to analyse the meaning of causal claims in the biomedical sciences. In order to explain how my definitions follow from Giere's work, I first present the core of his comparative model. It consists of the following definitions:

> C is a positive causal factor for E in the population U whenever $\mathbf{P}_X(E)$ *is greater than* $\mathbf{P}_K(E)$.
> C is a negative causal factor for E in the population U whenever $\mathbf{P}_X(E)$ *is less than* $\mathbf{P}_K(E)$.
> C is causally irrelevant for E in the population U whenever $\mathbf{P}_X(E)$ *is equal to* $\mathbf{P}_K(E)$.
> (Giere 1997, p.204)

Most of Giere's examples come from the biomedical sciences. So the population U mostly is a subclass of human beings, e.g. all Americans or all women in Germany. Giere considers only binary variables: C is a variable with two values (C and not-C); the same for E (values E and not-E). This is an important difference with my definitions, to which I come back below. For Giere, X is the hypothetical population which is obtained by changing, for every member of U that exhibits the value not-C, the value into C. K is the

[13] Note that "clean" here is to be interpreted strictly, as the state in which there are no substances present on the surface of the cable inside the sleeve. A lubricated cable has greasy material on its surface and thus is not clean in this sense.

analogous hypothetical population in which all individuals that exhibit C are changed into not-C. $\mathbf{P}_X(E)$ and $\mathbf{P}_K(E)$ are the probability of E in respectively X and K. Probabilities are defined as relative frequencies (Giere takes U to be finite, i.e. causal claims are about finite populations).

Let me give an example. If someone claims that smoking (**C**) is a positive causal factor for lung cancer (**E**) in the Belgian population (U), this amounts to claiming that if every inhabitant of Belgium were forced to smoke there would be more lung cancers in Belgium than if everyone were forbidden to smoke. Conversely for the claim that smoking is a negative causal factor. Causal irrelevance is a relation between variables (represented in bold) rather than a relation between values of a variable (like the first two relations). If we claim that "smoking behaviour" (**C**) is causally irrelevant for "the incidence of lung cancer" (**E**) this means that we believe that in the two hypothetical populations the incidence of lung cancer is equally high.

In my definitions, I will preserve the idea of hypothetical populations but formulate it in terms of hypothetical sets of setups, to accommodate my examples. I will also preserve the notation Giere uses for the relevant probabilities that are compared in the definitions: $\mathbf{P}_X(E)$ and $\mathbf{P}_K(E)$. Note that $\mathbf{P}_X(E)$ should not be confused with $\mathbf{P}_U(E|C)$. The latter is the relative frequency of E in the subclass C of the *actual* population U, while the former is the relative frequency of E in the *hypothetical* population X (which is defined starting from U but certainly not identical to U or to U∩C). In the smoking example, $\mathbf{P}_U(E|C)$ would be the relative frequency of lung cancer in the subclass of actual smokers in the real population of Belgium. $\mathbf{P}_X(E)$ is the relative frequency of lung cancer in the hypothetical population in which every inhabitant of Belgium were forced to smoke. Put differently: a difference between $\mathbf{P}_U(E|C)$ and $\mathbf{P}_U(E|\neg C)$ entails that the variables are correlated, but not that smoking is a cause of lung cancer. In order to have causation we need something else: a difference between $\mathbf{P}_X(E)$ and $\mathbf{P}_K(E)$.

To avoid misunderstanding, I provide some further clarification regarding the hypothetical populations. The idea underlying X and K is that they differ only with respect to the value of **C** and the values of all variables causally downstream of **C**. This is related to the way we 'change' the value of **C** in order to obtain the hypothetical populations. Giere does not specify what this change entails, but it can be defined as a surgical change following Jim Woodward (2003). Bear in mind that my account is a pragmatic one, focused on using causal knowledge. So this notion is primarily useful because reasoning about hypothetical populations and hypothetical changes to these populations can help understand the cases I am talking about.

Now that I have covered Giere's comparative model, I can present my definitions for the TPSIs and explain how they are related to Giere's.

2.3.2 Positive and negative causal factors

The first definition I need for my analysis is of a Positive Causal Factor (PCF):

(D_1) **PCF (Positive Causal Factor)**: C (as opposed to C*) is a positive causal factor for E in the collection of setups U if and only if $\mathbf{P}_X(E)$ is greater than $\mathbf{P}_K(E)$.

C and C* are mutually exclusive but not necessarily jointly exhaustive (for instance: a clean cable as opposed to a lubricated cable; there are other possibilities such as a gritty cable). The collection of setups (U) will consist of all individual setups that belong to a certain type (S). X is the hypothetical collection of setups which is obtained by changing, for every individual setup in U that does not exhibit the value C, the value into C. K is the analogous hypothetical collection of setups in which all individual setups that do not exhibit C* are changed into C*. $\mathbf{P}_X(E)$ and $\mathbf{P}_K(E)$ are the probability of E in X and K respectively.

An example might clarify this. If we claim that water in the fuel system (C) as opposed to only fuel in the fuel system (C*) is a positive causal factor for a sputtering engine (E) in the physical setups of U (setups of type S_1), this amounts to claiming that if we poured water into every fuel system that is part of a physical setup in collection U there would be more sputtering engines than if every fuel system of the setups in U was completely filled with fuel.

The second definition (Negative Causal Factor) is the negative counterpart of the first:

(D_2) **NCF (Negative Causal Factor)**: C (as opposed to C*) is a negative causal factor for E in the collection of setups U if an only if $\mathbf{P}_X(E)$ is less than $\mathbf{P}_K(E)$.

These definitions, as mentioned, characterise a rather weak notion of causation. To capture the no harm requirement and the ideal of maximal assistance, I need stronger notions. So my four remaining definitions (section 2.4 and 2.5) are reinforcements of D_1. It is possible to construct reinforcements of the above definition of negative causal factorhood. However, since I do not need these definitions for the current analysis, I leave this up to the reader.

Though my definitions of PCF and NCF are inspired by Giere's, there are some differences:

(a) I use "if and only if" while Giere uses "whenever" which is a conditional in one direction only. Because of the complementarity of Giere's three definitions (i.e. the fact that the three relations are jointly exhaustive and mutually exclusive) my biconditional

formulation is in fact equivalent to the original formulation. Because I will focus on positive causal factorhood, it is more convenient to use a biconditional formulation, since this avoids confusion when one definition is used in isolation from the other two.[14]
(b) Giere refers to populations, which is due to his interest in biomedical sciences, while my definitions are phrased in terms of collections of setups. The latter is of course due to my interest in technical problem solving instructions. Note that manuals often tell you the relevant collection of setups, by specifying the model numbers etc. So where Giere presupposes some sense of homogeneity in a population, manuals often are specific about their domain.[15]
(c) The third difference is philosophically the most important. While Giere only considers binary variables, my definitions allow for non-binary ones. So my take on variables is more general than Giere's. Recall the example at the end of in Section 2.1, for which the following claims may be true:

> The cable being clean (as opposed to lubricated) is a positive causal factor of braking difficulties.
>
> The cable being clean (as opposed to gritty) is not a positive causal factor of braking difficulties.

Here C refers to "clean". Clearly, there are two substates of not-C: "lubricated" and "gritty", and depending on which substate is seen as the alternative, the truth value of the causal claim differs.

2.3.3 Properties of PCF and NCF

It is immediately clear that the definitions of PCF and NCF have the characteristics I (the relevant type of physical setup is mentioned) , II (purely consequential definition) and IV (reference to the contrast cause) mentioned in Section 2.2.2. What about III (clarify why the causal knowledge is useful)? As we have seen, an important feature of Giere's

[14] If you believe that C is a positive causal factor for E in a population U, you cannot believe at the same time that C is a negative causal factor for E in U. Then (according to Giere's one-sided definition of negative causal factorhood) you have also to reject that $\mathbf{P}_X(E) < \mathbf{P}_K(E)$. A similar line of reasoning based on the definition of causal irrelevance leads to the rejection of $\mathbf{P}_X(E) = \mathbf{P}_K(E)$. Hence, you are forced to accept that $\mathbf{P}_X(E) > \mathbf{P}_K(E)$: this is the only option left. In this way it can be shown that Giere's definitions that use "whenever" jointly entail that in each case the other direction is also valid.

[15] Determining the domain of causal claims, however, is not easy. I will discuss this complication in chapter 4.

original model is that he defines causation in terms of what would happen in two hypothetical populations. In this way the policy relevance of biomedical causal claims becomes clear. Why should policy makers want causal knowledge? The hypothetical populations X and K correspond to populations a policy maker may create by means of some direct intervention (e.g. a ban on smoking, a mandatory inoculation, ...).[16] This feature is preserved in my definitions: they define causation in terms of what would happen in hypothetical sets of setups. Contrary to the policy maker, the TPSI user does not want to intervene on the entire set of setups. He is interested in the individual level: he wants to fix his specific setup, e.g. (a part of) his car. Knowledge about positive causal factors can still guide him in this quest. A person following the TPSI may create a member of the hypothetical set X, by means of some direct intervention (e.g. removing the water from the fuel tank). This is also why it is important to specify the considered alternative cause: it gives you more fine-grained information on which you can base your intervention. In the case of braking difficulties, the contrasts tell you that pouring dirty oil or dirty water in your cable will not solve the problem, since doing so will deposit grit in the cable.

I can now also explain why I used Giere's comparative model to base my definitions on, and no other, more discussed accounts of contextual causation like the one by Menzies (2007) or Hitchcock (1996). Giere's model has precedence, it was published in 1979. The definitions based on his model can also be a guide to interpret the biomedical and social examples in this dissertation (see chapters 3 and 4). But more importantly, Giere's idea of hypothetical populations fits really well with my cases. In general, the manuals tell you the set of setups their instructions hold for, by specifying what models they work for (e.g. Corvettes between 1963 and 1983 (Chilton 1986)) or by having you perform a diagnostic action and then specifying remedy instructions depending on the outcome of the actions. And the hypothetical changing of features also fits with my cases, since they deal with interventions. The manuals specify which changes can be made by the interventions that they prescribe. In a sense, they also tell you which changes cannot be made, by not including these interventions. This latter point is connected to the social embeddedness of manuals (see 2.7.2).

[16] Of course, there are often ethical and practical limitations here.

2.3.4 Implications of the efficiency requirement

In example E_4, the instruction is to drain the tank and blow out the fuel lines with compressed air. Suppose we find this instruction in a manual, while on the other hand we have good reasons to believe that the following claim is false:

> Containing only fuel in the fuel tank (C) as opposed to fuel and water (C*) is a positive causal factor for a normally running engine, in collections of setups of type S_1.

In such case, the manual proposes a useless intervention and thus the efficiency requirement is violated. In order for the instruction "drain the tank" to be suitable for inclusion (given the aims of the manual) this claim about positive causal factorhood must be true.

Similarly, in E_5 the instruction to recharge the battery is suitable for inclusion only if the following claim is true:

(C_7) The battery being fully charged (C) as opposed to only partially charged (C*) is a positive causal factor for a starter motor turning the engine at normal speed, in collections of setups of type S_3.

Where S_3 refers to:

(S_3) (1) the battery (with as variable feature whether it is charged) and (2) the starter motor (with as variable feature the speed with which it turns), in all cars.

Some remedy instructions rely on negative causal factorhood. Consider for example the instruction in E_1 to clean the air filter. This is suitable for inclusion if the following is true:

> A clean air filter (C) as opposed to a dirty one (C*) is a negative causal factor for excessive fuel consumption, in collections of setups of some type S_i.

Another example is $E_{1'}$. There, the instruction to adjust tyre pressure relies on the truth of the following claim:

> Properly pressured tyres (C) as opposed to underinflated tyres (C*) is a negative causal factor for excessive fuel consumption, in collections of setups of some type S_j.

These four examples support the thesis that, in order to be suitable for inclusion in a manual that respects the efficiency requirement, remedy instructions need to be backed

up by a true positive causal factor claim (where "positive causal factor" is defined as in D_1) or a true negative causal factor claim (where "negative causal factor" is defined as in D_2). If these claims are false, the remedy instructions are inadequate. And clearly, manuals tend to respect this requirement – for good reasons.

The constraint that is imposed here on remedy instructions is rather weak, because the definitions of PCF and NCF are not very demanding. If we are only looking to perform interventions that have some shot at working, this is enough. Yet, in the context of remedy instructions, there is a reason to look for stronger constraints, namely the no harm requirement. The truth of type level claims as defined by PCF and NCF is compatible with adverse effects in certain subsets of the population or the set of physical setups. Sticking to the engine examples: the truth of the positive or negative causal factorhood that backs up a remedy instruction is compatible with a situation in which you cause serious damage to your car's engine by executing the instruction. If we want to ensure that our intervention does not cause more damage, we (also) need a different property to hold for the causal relation than merely PCF. This is the main topic of Section 2.4.

2.4 The No Harm Requirement

I started with definitions of PCF and NCF because they are rather weak. Stronger definitions (that correspond to crucial ideas of Eells and Mackie) can be obtained by adding constraints. In this section I will use an important idea of Eells, viz. context unanimity, to clarify how the no harm requirement can be satisfied. In 2.4.1 I will explain what context unanimity is and why it is not present in Giere's model. In 2.4.2. I will incorporate the idea of context unanimity in my analysis by means of two definitions which are reinforcements of D_1. In 2.4.3 and 2.4.4 I will relate these two definitions to the no harm requirement.

2.4.1 Average effect versus context unanimity definitions of causation

Giere's original comparative model and my adaptation can be characterised as "average effect" definitions of causation in sets of physical setups and biological populations. Let me clarify what this means by looking at an example from the biomedical sciences. Consider a dangerous virus, which threatens a population of humans (H). Some people

are immune to the disease (I), but there is no way to find out who is and who is not. It is possible to vaccinate people before they become sick (V). I assume the following probabilities in the hypothetical populations (S stands for survival):

$$\boldsymbol{P}_V(S \mid I) = 0.9$$
$$\boldsymbol{P}_V(S \mid \neg I) = 0.8$$
$$\boldsymbol{P}_{\neg V}(S \mid I) = 1$$
$$\boldsymbol{P}_{\neg V}(S \mid \neg I) = 0$$

Furthermore, I assume that 50% of the population is immune, so we also have:

$$\boldsymbol{P}_V(S) = 0.85 \qquad (0.8{\times}0.5 + 0.9{\times}0.5)$$
$$\boldsymbol{P}_{\neg V}(S) = 0.5 \qquad (1{\times}0.5 + 0{\times}0.5)$$

Note that in the subpopulation of people that are immune (I), vaccination is a negative causal factor: 10% of this subpopulation would not survive vaccination. In subpopulation $\neg I$ and in H as a whole, vaccination is a positive causal factor.

How is this possible? In I there is a group of people (10% of I) whose residual state is such that they die if vaccinated. For the others, vaccination is causally irrelevant at the individual level. Combined, this gives a negative causal relevance at the level of subpopulation I. In subpopulation $\neg I$ we have a large group (80%) whose residual state is such that vaccination is positively causally relevant at the individual level. For the others, it is irrelevant (their residual state is such that they die anyway). The combination of this gives positive causal relevance at the level of subpopulation $\neg I$. The population H contains a group of 5% (10% of the 50% immune) for whom vaccination has negative causal relevance at the individual level. It also contains a group of 40% (80% of the 50% non-immune) for whom vaccination has positive causal relevance at the individual level. For the others, vaccination is causally irrelevant at the individual level (vaccination makes no difference for them: they survive anyway because they are immune, or they die anyway because the vaccination does not work for them). Because the group with positive relevance is larger (40% as compared to 5%) the result is a positive causal relevance at the level of population H.

The vaccination example illustrates that, according to Giere's definitions, causal relevance can be reversed or annihilated in subpopulations: if C is a positive causal factor for E in population U, it can be a negative causal factor or be causally irrelevant in

subpopulations of U[17]. The same holds for negative causal factors. Theories of causation which have this property are called "average effect theories": whether there is a causal relation in a population depends – according to these theories – on the average effect in the population, no matter what happens in its subpopulations (Weber 2009, p.283).

An alternative to average effect theories are the so-called "context unanimity theories". The first context unanimity theory can be found in (Cartwright 1979). A more recent version can be found in Eells (1991). In chapter 2 of his book, Eells gives the following example:

> To use an example of Cartwright's (1979), ingesting an acid poison (X) is causally positive for death (Y) when no alkali poison has been ingested (~F), but when an alkali poison has been ingested (F), the ingestion of an acid poison is causally negative for death. I will argue that in a case like this it is best to deny that X is a positive causal factor for Y, even if, overall (for the population as a whole), the probability of death when an acid poison has been ingested is greater than the probability of death when no acid poison has been ingested (that is, even if Pr(Y/X) > Pr(Y/~X)). I will argue that it is best in this case to say that X is causally mixed for Y, and despite the overall or average probability increase, X is nevertheless not a positive causal factor for Y in the population as a whole. (Eells 1991, p.58)

The characteristic property of causes in the sense of context unanimity theories is that the causal tendency cannot be reversed (from positive to negative) or annihilated (from positive or negative to causally neutral) in a subpopulation. Note that I only borrow the idea of context-unanimity from Eells and not the remainder of his theory. This is because Eells considers *actual* probability distributions in defining causation, instead of the hypothetical ones Giere uses. This is already clear from the quote above: he defines a causal factor and the property of being causally mixed in terms of the actual probability of the occurrence of the effect given the presence or absence of the cause (Pr(Y/X) and Pr(Y/~X) respectively). Because I want to preserve the advantages of Giere's definitions in terms of hypothetical populations, I only borrow Eells' idea of context-unanimity.

[17] C can even be negative causal factor for E in *every* subpopulation of U. This is referred to as Simpson's paradox" (Malinas and Bigelow 2016).

2.4.2 Two additional definitions

The idea of context unanimity can give rise to two reinforcements of PCF: PCF-WU and PCF-SU (where WU stands for weak unanimity and SU for strong unanimity):

(D_3) **PCF-WU (Positive Causal Factor – Weak Unanimity)**: C (as opposed to C*) is a weakly unanimous positive causal factor for E in the collection of setups U if and only if
(1) $\mathbf{P}_X(E)$ is greater than $\mathbf{P}_K(E)$; and
(2) there are no subgroups S of U for which $\mathbf{P}_{X'}(E)$ is less than $\mathbf{P}_{K'}(E)$.

X' and K' are defined in the same way as X and K above, but starting from the subset of setups S instead of the whole set of setups of a certain type U.

(D_4) **PCF-SU (Positive Causal Factor – Strong Unanimity)**: C (as opposed to C*) is a strongly unanimous positive causal factor for E in the collection of setups U if and only if
(1) $\mathbf{P}_X(E)$ is greater than $\mathbf{P}_K(E)$; and
(2) there is no subgroup S of U for which $\mathbf{P}_{X'}(E)$ is less than or equal to $\mathbf{P}_{K'}(E)$.

The difference between the two definitions is that weak context unanimity allows that a causal tendency in the whole population is annihilated in a subpopulation. It only prohibits tendency reversal (viz. from a positive causal tendency to a negative one or vice versa). Strong context unanimity prohibits both tendency annihilation *and* tendency reversal.

2.4.3 Remedy claims and weak context unanimity

There is a connection between the no harm requirement and weak context unanimity. In order to argue for this, imagine a hypothetical sloppy manual that contains the following TPSI:

> Problem: engine sputters and then stops running.
> Diagnosis: empty fuel tank.
> Remedy: put gasoline in the fuel tank.

The sloppiness consists in the fact that this instruction is given for different varieties of the specific car type (let me call that S4). Specifically, imagine that this instruction is included in the manuals of both the diesel versions of S4 and the gasoline ones. Let us

assume that 80% of the cars that belong to type S4 have a gasoline engine. Then the following causal claim is correct:

> The fuel tank containing gasoline (C) as opposed to diesel (C*) is a positive causal factor for a normally running engine, in collections of setups of type S_4.

Indeed, if all fuel tanks in S_4 cars contain diesel, only 20% runs normally. If all fuel tanks contain gasoline, 80% runs normally. Let S5 be all the cars of type S4 with a diesel engine. Then the following claim is true:

> The fuel tank containing gasoline (C) as opposed to diesel (C*) is a negative causal factor for a normally running engine, in collections of setups of type S_5.

So S_4 violates weak context unanimity. The following claim is false:

> The fuel tank containing gasoline (C) as opposed to diesel (C*) is a weakly unanimous positive causal factor for a normally running engine, in collections of setups of type S_4.

The fact that there is no weak context unanimity entails that there are contexts (or subgroups of setups - in casu: cars with diesel engines) where the action “put gasoline in the fuel tank” worsens the problem rather than helping to solve it. This is why the proposed remedy “put gasoline in the fuel tank” is not suitable for manuals that cover both diesel and gasoline versions. In reality, manuals are rather specific in mentioning the artefact types or models for which they hold, as I mentioned in 2.3.3. This allows them to ensure that sloppy TPSIs like the one above do not occur. Manuals tend to respect the no harm requirement.

2.4.4 Remedy claims and strong context unanimity

So weak unanimity is a desirable feature for causal claims underpinning remedy instructions. Should we go even further and expect strong unanimity? This would be a step too far. This can be seen from two examples in which there is no strong context unanimity. The first example is C_7 from 2.3.4:

(C_7) The battery being fully charged (C) as opposed to only partially (C*) is a positive causal factor for a starter motor turning the engine at normal speed, in collections of setups of type S_3.

There are subsets in which the positive difference is annihilated, e.g. in the set of cars with heavily corroded battery terminals. A fully charged battery makes no difference for the behaviour of the engine in setups of that type. So there is no strong context unanimity in this case. The following claim is false:

> The battery being fully charged (C) as opposed to only partially charged (C*) is a strongly unanimous positive causal factor for a starter motor turning the engine at normal speed, in collections of setups of type S_3.

A similar observation can be made with respect to the following example (based on C_6 above):

($C_{6'}$) Containing only fuel in the fuel tank (C) as opposed to fuel and water (C*) is a positive causal factor for a normally running engine, in collections of setups of type S_1.

Again there are subsets in which this positive difference is annihilated, e.g. in the set of cars with defective fuel pumps. A properly filled fuel tank makes no difference for the behaviour of the engine in setups of that type. This means that, again, there is no strong context unanimity. The following claim is false:

> Containing only fuel in the fuel tank (C) as opposed to fuel and water (C*) is a strongly unanimous positive causal factor for a normally running engine, in collections of setups of type S_1.

The crucial question now is: is this absence of strong unanimity a problem? No, because no harm can be done if there is weak context unanimity. Remember that the manuals often specify the artefacts they are suited for. So you can easily determine whether your artefact is part of the domain, and correspondingly, weak unanimity will hold. When there is weak context unanimity, there are no contexts in which the remedy instruction worsens the problem. Moreover, TPSIs are typically part of a larger set of instructions. If one instruction does not lead to a functioning artefact, there often is a follow-up instruction that prescribes another action aimed at solving the problem. In this way, a TPSI not need not be self-contained, but functions as part of a bigger whole.

So depending on whether we want our interventions to (1) not be useless or (2) not be useless and do no harm, we need different causal knowledge to back up the intervention. This already shows that the meaning of a causal claim is not straightforward and neither is determining which properties are needed to warrant an intervention. Yet we still do not have any guarantee that our intervention will solve the problem. What causal knowledge do we need to ensure that it will? This is the topic of the next section.

2.5 The Maximal Assistance Ideal

So far, I have specified how the efficiency requirement and the no harm requirement influence the causal knowledge needed to warrant remedy claims in repair manuals. In this section, I turn to the maximal assistance ideal. Remedy claims that are based on true causal claims in the sense of PCF-WU do not guarantee that the problem is solved. In order to guarantee a certain outcome, we need a cause that is *sufficient* to bring the outcome about. So maybe remedy claims should be based on knowledge about sufficient causes? In 2.5.1 I will argue that this is not a good idea. In 2.5.2 and 2.5.3 I will develop an alternative based on Mackie's concept of INUS condition.

2.5.1 Sufficient causation?

In order to guarantee that the problem is solved if the prescribed action is performed, we need a stronger causal relation. A concept of causation that captures this stronger causal relation can easily be obtained by adding a sufficiency clause to definition D_1 of PCF. This results in the following definition:

(D_5) **SC (Sufficient Cause)**: C (as opposed to C*) is a sufficient cause for E in the set of physical setups U if and only if
(1) $\mathbf{P}_X(E)$ is greater than $\mathbf{P}_K(E)$; and
(2) $\mathbf{P}_X(E)=1$.

Note that, according to this definition, a sufficient cause is always a positive causal factor. The sufficiency condition also implies that there is weak (and even strong) context unanimity. If $\mathbf{P}_X(E)=1$ as required by the second clause in D_5, it will also be the case that $\mathbf{P}_{X'}(E)=1$ for all subsets of U. Put differently: if all objects in X have a property E, then it is also the case that, for all subsets of X, all their members have E.

So does this mean that adequate repair manuals should include only remedy claims that are based on true sufficient cause claims? This would imply that most of the current manuals are significantly substandard. That seems highly unlikely. And indeed, expecting only sufficient causal claims is way too demanding. The reason is that there are hardly any true claims of this kind available. For instance, the following claims are all *false*:

> The battery being fully charged (C) as opposed to only partially charged (C*) is a *sufficient* cause for a starter motor turning the engine at normal speed, in collections of setups of type S_3.

Containing only fuel in the fuel tank (C) as opposed to fuel and water (C*) is a *sufficient* cause for a normally running engine, in collections of setups of type S_1.

These claims are false because the problems that are to be solved may have multiple causes. In the first case, the battery terminal may be corroded. The second claim is false because the fuel pump may be defective.

Because of this scarcity, the requirement that remedy claims are to be based on sufficient causation relations as defined by D_5, is untenable: repair manuals would hardly contain any instructions. This would be at odds with the maximal assistance ideal.

So it looks as if we have to allow remedy claims to be based on non-deterministic causal relations. But given the maximal assistance ideal it is useful to investigate whether some weaker alternative is possible. Inspired by John Mackie's work, I think there is such an alternative, and I call it "sufficiency in maximally normal contexts". In Section 2.5.2 I will clarify what I mean by this and why it is a desirable property of TPSIs. In section 2.5.3 I will show that the idea can be captured by adding Mackie's concept of INUS condition to the definition of PCF.

2.5.2 Sufficiency in maximally normal contexts

The issue I want to bring up here can be illustrated by means of the water-in-the-fuel-system example. Let me compare two TPSIs. The first is:

Problem: engine sputters and fails to start.
Diagnosis: water in fuel supply system.
Remedy: drain the fuel tank.

The second is:

Problem: engine sputters and fails to start.
Diagnosis: water in fuel supply system.
Remedy: drain the fuel tank and blow out the fuel lines with compressed air.

Which instruction is the best one? In general, draining the fuel tank is not sufficient for solving the problem because there may also be water in the fuel lines. So the second TPSI is more adequate. This indicates that we want the remedies to be complete in some sense: they have to be sufficient for solving the problem provided that there is nothing wrong with the artefact on top of what is stated in the diagnosis. In other words, the remedies have to be complete relative to the diagnosis that is part of the TPSI. This is what I call the "maximally normal context": the context where all components – except

the one addressed in the diagnostic claim – of the artefact function normally (i.e. as conceived by the designers, and as they can legitimately be expected to function in a newly produced artefact that has passed the manufacturer's quality control).

2.5.3 Mackie-causation and its application to TPSIs

Mackie's theory of causation is situated at the token level: it is about causal relations between particular events. I am dealing here with causal relations between variables, so I will adapt it a bit. I first give a brief presentation of Mackie's account and then clarify how I propose to use the crucial INUS concept at the type level in combination with my definition of PCF.

In *The Cement of the Universe* Mackie claims that a cause is always (at least)[18] an INUS condition for its effect (1980, p.64). He defines INUS conditions as follows:

> […] an insufficient but non-redundant part of an unnecessary but sufficient condition: it will be convenient to call this (using the first letters of the italicized words) an inus condition. (Mackie 1980, p.62)

So in Mackie's view, a cause in itself need not be not sufficient for its effect, but combined with a set of other factors, it is. Furthermore, a cause is typically not necessary for its effect, since multiple sets of factors can produce the same effect. Consider one of his examples:

> The short-circuit caused the fire.

As Mackie points out, the short-circuit (c) in itself is not sufficient for a given fire (e): you also need oxygen and combustible material (A). The factors in (A) are themselves also not sufficient for the fire (e), so the short-circuit is a non-redundant part of the condition (A+c). Combined, the short-circuit and the factors in A are sufficient for the fire. Yet (A+c) is not necessary for a fire to happen, a fire can also occur because of (A+ lightning), (A+ a lit match), etc.

By adding Mackie's crucial INUS idea as a constraint to the definition of PCF and linking it to maximally normal contexts, it can be incorporated in my type level analysis. This results in my sixth definition:

[18] For some more recent refinements of Mackie's INUS account, see (Baumgartner 2008) and (Beirlaen, Leuridan, and Van De Putte 2016).

(D_6) **MC (Mackie Cause)**: C (as opposed to C*) is a Mackie cause for E in the collection of setups U if and only if
(1) $\mathbf{P}_X(E)$ is greater than $\mathbf{P}_K(E)$; and
(2) if A characterises the maximally normal context, it is the case that (a) all members of U that have A and C, also have E, and (b) not all A have E.

Let me give an example based on the following TPSI:

> Problem: starter motor turns engine slowly.
> Diagnosis: partially discharged battery.
> Remedy: recharge.

Let C stand for "fully charged battery" and E for "starter motor turns engine rapidly". C is not a sufficient cause for E, because there may be other malfunctioning components (e.g. corroded battery terminals, bad starter motor, …). But C in combination with A (where A is a description of the maximally normal context, including a.o. the proviso that the battery terminals are not corroded and that the starter motor is not broken) is sufficient for E.

Or consider the radio example (E_8). We can isolate the following TPSI:

> Problem: the sound produced by the speaker is distorted.
> Diagnosis: the cone of the loudspeaker is badly torn or warped.
> Remedy: replace the cone with an intact one or recone the old one.

According to my analysis at the very minimum the following causal claim has to be true to make the remedy claim warranted:

(C_8) An unwarped, untorn cone (C) instead of a warped or torn cone (C*) is a weakly unanimous positive causal factor for a loudspeaker that produces a clear sound (E) for the setups of type S_4.

where

(S_4) (1) The cone (with as variable features whether it is (a) torn and (b) warped) and (2) the loudspeaker (with as variable feature whether it produces distorted sound), in all radios with conic loudspeakers.

And ideally, MC is also satisfied. Applied to this example, MC comes down to requiring that if the radio is functioning properly on all fronts except the torn or warped cone, replacing the cone (without any further additional actions) will ensure that the sound is no longer distorted. So if the radio does not suffer from other problems (e.g. dirty tube

socket connections, open capacitors,…) that can hinder its functioning, the ideal outcome is that, after the prescribed intervention, the radio functions properly.

So adding the maximal assistance ideal as a criterion requires an even stronger causal relation to underpin remedy claims, and it can be captured via the definition of Mackie Cause.

2.6 Synthesis and further reflections

Before I turn to the third complicating factor, I want to summarise how the analysis developed in sections 2.3 – 2.5 presents a second way in which, contrary to what Norton assumes, ascribing causal relations to phenomena is complicated. This will allow me to further develop my argument to show that the way we make physical causal claims, and use physical causal knowledge, is philosophically interesting and requires attention. I will also reflect a bit on some particularities of the presented definitions. Let me first summarise the content of the sections.

In the three previous sections, I have performed a stepwise analysis of how the success criteria for instructions in manuals determine the nature of the causal knowledge that is required to underpin remedy claims. From manuals that want to satisfy the efficiency requirement and the no harm requirement, it is legitimate to expect underlying causal relations in the following sense:

(D_3) **PCF-WU (Positive Causal Factor – Weak Unanimity)**: C (as opposed to C*) is a weakly unanimous positive causal factor for E in the collection of setups U if an only if
(1) $\mathbf{P}_X(E)$ is greater than $\mathbf{P}_K(E)$; and
(2) there are no subgroups S of U for which $\mathbf{P}_{X'}(E)$ is smaller than $\mathbf{P}_{K'}(E)$.

I have discussed this in 2.3 and 2.4. In Section 2.5 I reflected on the maximal assistance ideal. I argued that it motivates a preference for sufficiency in maximally normal contexts:

(D_6) **MC (Mackie Cause)**: C (as opposed to C*) is a Mackie cause for E in the collection of setups U if and only if
(1) $\mathbf{P}_X(E)$ is greater than $\mathbf{P}_K(E)$; and
(2) if A characterises the maximally normal context, it is the case that (a) all members of U that have A and C, also have E, and (b) not all A have E.

In other words: one who agrees with the three criteria I formulated in the introduction can legitimately expect positive causal factors that are weakly context-unanimous and legitimately prefer Mackie causes. Requiring only positive causal factorhood (PCF) as defined in D_1 is not enough. In order to satisfy the criteria, knowledge that expresses a stronger causal relation is required.

Note that both concepts are needed. It is easy to see that, if a purported cause satisfies PCF-WU, it may fail to satisfy MC. Though less obvious, the opposite may also be the case; a state (C) that is sufficient for solving a problem in the maximally normal context (A), may be harmful in other contexts (i.e. if there are other problems with the artefact besides the one mentioned in the diagnosis).

In light of the maximal assistance ideal, stronger requirements or preferences are not optimal. As I argued in Section 2.5.1, including only remedy claims based on sufficient causation (as defined in D_5) is not the most favourable strategy: the manuals would be almost empty. Artefacts are such complex systems that determining a sufficient cause for a problem is extremely hard, both to formulate and to diagnose. A similar argument can be made with respect to strong context-unanimity (as defined in D_4). In Section 2.4.4 I have argued already that strong context-unanimity it is not necessary in order to avoid harm. However, we can go one step further: if composers of repair manuals would restrict themselves to instructions based on strongly unanimous causal relations, they would have to leave out many adequate remedies. For instance, they would have to leave out remedies such as “recharge the battery” or “drain the fuel tank” referring to C_7 and C_6 from 2.4.4. Yet, as I argued, performing the action is never harmful and it has a significant chance of fixing the problem. So it is not surprising that these remedies are in fact included in the manual.

What this analysis shows regarding the difficulties for ascribing causal relations to physical phenomena is that, depending on the demands we have for our intervention (represented by the three criteria), we need different causal knowledge to warrant it. So when we are interested in interventions, ascribing a causal relation to a phenomenon requires, besides information from the context (viz. the first complicating factor), information about the specific demands for the intervention. Otherwise, the causal claim is not adequate to base our intervention on. I hope to have shown that determining when and if a causal claim is of the right kind to warrant our intervention, is not a straightforward matter.

In the next section, I turn to the third complicating factor. So far, my analysis has been limited to the specific instructions and the specific malfunctioning they attempt to remedy. However, the situation is more complex. In the next section, I will discuss the broader physical and social embeddedness of artefacts – the third complicating factor.

2.7 The Third Complicating Factor

As I described in 1.5.3, artefacts are complex arrangements of interacting parts. So the malfunctioning parts are embedded in a bigger artefact. Those artefacts are also embedded in a broader physical and social environment. Though this is not necessarily reflected in the remedy claims, these environments can also influence the effects of our interventions. I will argue that these environments present the third complicating factor for ascribing causal relations to phenomena when we are interested in using that causal knowledge for intervening in the phenomenon. The reflections in this section are inspired by an important topic in the philosophy of technology that I discussed in 1.5.7: a comprehensible account of technology is aware of the social and physical embeddedness of artefacts. The broader physical and social environment are even less reflected upon in the philosophical literature on physics than the other complicating factors, but are nevertheless important and intriguing when we want to understand how physical causal knowledge is used.

2.7.1 The physical environment

In discussing the first and second complicating factor, I have focused on the specific TPSI in isolation. Philosophers of physics, like Norton, often do the same with examples: the phenomenon in question is analysed in isolation from anything else and therefore looks quite uncomplicated. This is related to the amount of control we have in laboratories: we can apply proper shielding to limit the influence of an on the broader environment as much as possible. However, in reality, these phenomena are embedded in bigger physical and even social environments that are not easily shielded.

Something I have not reflected on so far is that interventions resulting from remedy instructions, if performed properly, should not create new problems in the artefact (the physical environment of the malfunctioning parts) either. For example, replacing the cone should not influence the display of the radio. If one wants to express this in relation to the definitions above, it would look something like this:

> An instruction that tells you to change C to C* is safe in the collection of setups U if and only if
> (1) $\mathbf{P}_X(E')$ is equal to or less than $\mathbf{P}_U(E')$; and
> (2) there are no subgroups S of U for which $\mathbf{P}_{X'}(E')$ is significantly greater than $\mathbf{P}_U(E')$.

Where E' is an unwanted effect, U is now the collection of specific setups that are *suffering from the diagnosis as stated in the TPSI,* X and X' are defined as before and $\mathbf{P}_U(E')$ is the chance of the unwanted effect occurring in the collection of setups that suffer from the diagnosis, without us performing any intervention. The idea behind this is that a prescribed action cannot significantly increase the chance of an unwanted effect occurring, compared to the chance it has of occurring in the malfunctioning artefact when you do not intervene. This unwanted effect could be another malfunction that was not yet present or even some kind of harm to the user. For example, if replacing the cone significantly raises the chance of the antenna failing, compared to the chance the antenna has of failing in the malfunctioning radio without performing the instruction, the instruction is not safe. This might look negligible, but given that artefacts often contain hazardous materials, like battery fluid, it is important to have some way of expressing that the prescribed interaction is safe, both for the other parts of the artefact and for the user. In a lot of contexts, there will be some kind of threshold to determine when the risk of failure is too great. This is especially a case of expert judgements. Summing up, the broader consequences of an intervention influence the outcome and are (rightfully) taken into account.

This is also related to the degree in which the artefact was designed modularly. The topic of modularity has been around for a while, and is now mainly discussed in philosophy of the life sciences and of the sciences of mind and brain. It refers to the fact that in many complex systems, not all parts are fully integrated. Instead,

> […] the parts of the system are grouped in such a way that strong interactions occur within each group or module, but parts belonging to different modules interact only weakly. (Krohs 2009, p.259)

Moreover, not all artefacts are designed equally modularly. As Ulrich Krohs (2009, pp.267-268) discusses, there are costs to modular designing regarding time and resources that need to be weighed. In order to assure that interventions do not harm any other parts of the artefact, we need knowledge of the degree of modularity and the boundaries of the modules. This puts even more demands on the causal knowledge required for intervening.

2.7.2 The social environment

Ensuring that no other parts of the artefact suffer harm for my cases still mainly refers to very proximate and clearly physical factors. However, artefacts and interventions in general are also embedded in a social environment. One way in which artefacts are

socially embedded is via the users of the artefacts (and the repair manuals). Resources and skills that are needed to perform the interventions may not be available to everyone. Users may ignore the social authority of the manuals. All of this influences the way the interventions will turn out. So the causal knowledge needed to warrant interventions involves more than merely knowledge of the physical artefact.

Repair manuals also function in a society where there are specified repair technicians. This influences what people expect of the manuals and what the manuals prescribe. Sometimes, the manuals tell you to take your malfunctioning artefact to a specialised technician. But for other problems, you would never use a manual and take them directly to a technician, or throw the artefact out. For example, if your bike was set on fire, you will take it to a technician. This is however not stated in manuals explicitly, but is part of how the manuals work in our society.

Similarly, what is often not mentioned in the manuals (though on occasion, they do) is the requirement of regular maintenance. However, the remedy instructions presuppose that the artefact has been maintained on a regular basis.

For the repair interventions to be successful, even when performed correctly, several other societal factors need to be in place. The radio, for example, needs electricity to work. If the user did not pay his bill, his electricity will be shut off. This too influences the functioning of the artefact. Similarly, radio pirates can jam the broadcast or an angry neighbour can sabotage the antenna.

All of these factors can be causal influences on the artefact and correspondingly, knowledge of these factors can be needed to warrant our interventions. In practice, it is impossible to incorporate all these influences. My point is not that they *need* to be taken into account to warrant interventions. My point is that when assessing causal relations, the broader physical and social factors are often presumed to be in place. Yet in reality, they are not easily controlled and complicate the evaluation of causal claims needed for interventions significantly.

Conclusion

I started this chapter with a quote from Norton, which indicates that he does not consider ascribing causal relations to phenomena to be an important philosophical topic. This assumption seems to underlie many debates in philosophy of physics (see 1.2). This chapter forms the first step in my argument to show the complexity of using physical causal knowledge and the importance of philosophical reflection on the topic. I discussed

three complicating factors that arise even when we consider rather easy and day-to-day contexts of using physical causal knowledge

The first complicating factor, I argued, is the fact that causal relations and more importantly, the validity of causal claims, depend on the context in the way I have specified, viz. on the (set of) physical setups you consider. In order to use causal knowledge for intervening, we need to ensure that the knowledge is applicable to the system we want to intervene in. I developed the tool of physical setup to capture this.

The second complicating factor I diagnosed, is that the required properties of the causal relation we need to warrant interventions depend on what we expect of the intervention. I argued that, in order for remedy claims to be adequate, they have to be based on causal relations that (i) hold in the world and (ii) have certain properties: positive or negative causal factorhood (D_1) with weak context unanimity (D_3). That these properties are required follows from two success criteria: the efficiency requirement and the no harm requirement. Positive and negative causal factorhood are explicated in definitions of PCF and of NCF, which are inspired by Giere's comparative model of causation. Weak context unanimity is explicated in a definition of PCF-WU and has its origins in the idea of context unanimity as put forward by Cartwright and Eells. Based on the maximal assistance ideal, I also argued that sufficiency in maximally normal contexts is desirable. I explicated it in the definition of MC (D_6) which is the result of adding Mackie's INUS idea to Giere's comparative model.

The third complicating factor is the social and physical embeddedness of the systems we intervene in. To ensure that our intervention does not harm other parts of the artefact, or even the users, we need causal information about the way the artefact or malfunctioning part is connected to the broader physical world. Moreover, the artefacts we intervene in are embedded in a social environment as well. These social environments can also influence the outcomes of our interventions. Therefore, knowledge of the way artefacts are socially embedded can also be relevant for warranting interventions. These broader environments are often neglected. However, while we can apply proper shielding in laboratory contexts, we cannot do this when intervening in reality. As announced, this chapter presented an argument for my first specific claim (I), viz. that we need a lot of specific physical causal information of the right kind, both of the artefact and of the physical and social context it functions in, to account for our successful creating, explaining, repairing, and maintaining of artefacts,

What I have also shown throughout this chapter, is that ascribing causal relations to physical phenomena in a way that can guide interventions, is not straightforward at all. In the case of repair manuals, optimising the content of a manual in light of these complicating factors, requires expert judgement. As I have argued, Mackie causes are desirable but there are other factors that may influence the decisions about which

remedy claims are included and how precise they should be. Since repair manuals target a non-expert, non-professional audience, feasibility may be such a factor. Manuals typically prescribe actions which are relatively easy to perform (like recharging a battery, cleaning, draining, blowing,….). Another trade-off may be specificity versus intelligibility: if made more specific, the success rate of an instruction may rise, but at the same time, it might make the instruction harder to understand. My analysis can therefore also help make explicit some of the criteria used by writers of repair manuals.

At the same time, I hope it shows some of the complexity of practices that we often take for granted. Repair manuals, bikes, cars, they are all very day-to-day and unprestigious topics. Yet they are riddled with reliable causal knowledge, and as I have shown, contain a lot of complexity below the surface.

The conclusions of this chapter are broadly relevant. As I discussed in chapter 1, philosophers of physics often focus on theory and idealisations of phenomena. As such, the context and broader physical and social environment are often neglected. However, when attempting to use physical causal knowledge, we need information about these factors. I will return to this topic in Chapter 4.

As I explained, this chapter forms the first step in my argument that shows the intriguing nature of using physical causal knowledge and the need for philosophical reflection on the topic (viz. aims (A) and (B)). The complexity of the cases I studied here was rather restricted: there is plenty of reason to trust the information contained in the manuals, since they are written by experts and have social authority. Moreover, the information is very connected to the interventions: manuals literally contain prescriptions regarding what you need to do, and they make explicit mention of their domain. This is not a situation we encounter a lot. We are often unsure whether our information is reliable, or whether it is applicable to the specific cases we are interested in, or we do not have the required knowledge to warrant the envisioned application. I started with the cases from this chapter, because they allowed me to discuss them in such a constrained way. In what follows, I investigate how using and producing causal knowledge becomes more and more complex as I focus on other contexts of application. In the next chapter, I investigate what happens when we do not have such a clear source of causal knowledge as the manuals provided.

References

Baumgartner, Michael. 2008. Regularity Theories Reassessed. *Philosophia* 36 (3):327-354.

Beirlaen, Mathieu, Bert Leuridan, and Frederik Van De Putte. 2016. A logic for the discovery of deterministic causal regularities. *Synthese*:1-33.

Carr, Joseph. 1990. *Old Time Radios! Restoration and Repair*. 1 edition ed: McGraw-Hill Education TAB.

Cartwright, Nancy. 1979. Causal laws and effective strategies. *Nous*:419-437.

Chilton. 1986. *Corvette, 1963-83*. 1 edition ed. Radnor, Pa: Chilton Book Company.

Considine, Glenn D., and Peter H. Kulik. 2008. *Van Nostrand's scientific encyclopedia*. Hoboken, N.J.: Wiley.

Eells, Ellery. 1991. *Probabilistic causality*. Vol. 1: Cambridge University Press.

Giere, Ronald. 1997. *Understanding Scientific Reasoning*. Forthworth: Harcourt Brace College Publishers.

Hitchcock, Christopher Read. 1996. Farewell to binary causation. *Canadian Journal of Philosophy* 26 (2):267-282.

Krohs, Ulrich. 2009. The cost of modularity. *Functions in Biological and Artificial Worlds: Comparative Philosophical Perspectives* 276:259-276.

Mackie, John L. 1980. *The Cement of the Universe*. 2nd ed: Oxford University Press.

Malinas, Gary, and John Bigelow. 2016. Simpson's Paradox. In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta: Metaphysics Research Lab, Stanford University.

Mead, John S., and A. K. Legg. 1997. *Toyota Carina E Service and Repair Manual*, *Haynes Service and Repair Manual Series*. Sparkford: Haynes Manuals Inc.

Menzies, Peter. 2007. Causation in Context. In *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, edited by H. Price and R. Corry: Oxford University Press.

Sidwells, Chris, and Richard Ballantine. 2004. *Bicycle Repair Manual*. New York: DK.

Steel, Daniel. 2007. *Across the Boundaries: Extrapolation in Biology and Social Science*: Oxford University Press.

Strasman, Peter. 1988. *Rover 213 & 216 Owners Workshop Manual*, *Haynes Service and Repair Manual Series*. Sparkford: Haynes Manuals Inc.

von Wright, Georg Henrik. 1971. *Explanation and Understanding*: Cornell University Press.

Weber, Erik. 2009. How Probabilistic Causation Can Account for the Use of Mechanistic Evidence. *International Studies in the Philosophy of Science* 23 (3):277-295.

Woodward, James. 2003. *Making Things Happen*: Oxford University Press.

Chapter 3 Mechanistic vs. Correlational evidence for physical causal explanations[1]

In the first chapter, I argued that we need physical causal information to intervene in the world and explain physical phenomena. The second chapter showed that, pace philosophers like Norton, (producing) the causal knowledge we need for interventions is not uncomplicated. In this chapter, I will focus on where we get this physical causal knowledge from. So in chapter 2 I argued that there are philosophical issues about the meaning of useful physical causal knowledge, here I shift to evidence. Specifically, I will investigate how we argue for physical causal claims used in *explanations* – the second way we use physical causal knowledge as I argued in chapter 1. Because the laws of physics are often symmetric, they do not suffice to argue for a specific causal claim. I call this an *evidential gap* between the laws and causal claims. This is the main issue related to evidence that I will discuss in this chapter. I will show that this gap can be bridged by *mechanistic evidence*. This concept has recently been discussed a lot in the context of the biomedical sciences. I will draw inspiration from this literature. However, certain methods for gathering correlational data that are frequently used in the biomedical sciences, provide correlational evidence that has different properties than the laws of physics. Because of this, the evidential gap in the biomedical sciences often differs from the one we encounter when arguing for physical causal claims. To make this clear, I will look at a case from the social sciences where the correlational evidence is less informative, and the evidential gap is more similar to the one for physical causal claims. This will give me the framework to both argue that mechanistic evidence can bridge the evidential gap in physics, and at the same time, reflect on the different evidential gaps that exist in these different disciplines. Specifically, the way we arrive at the laws or

[1] This chapter is based on a paper co-authored with Erik Weber.

regularities will determine the specific nature of the evidential gap and correspondingly, the role of mechanistic evidence.

This is related to the difference between the regularities in the biomedical and social sciences on the one hand, and the laws of physics on the other. Regarding physics, the analysis in this chapter will show that even for causal claims about setups that clearly fall under highly established laws of physics, we still need a significant amount of mechanistic information about the setup if we want to make and use causal claims about that setup. In this way, this chapter provides an argument for my second specific claim (II). This conclusion also contributes to realising my generic aims (A) and (B), by showing that there are interesting philosophical issues related to evidence we develop when producing explanatorily useful physical causal knowledge. Finally, this conclusion also contributes to my argument to show that the focus of philosophy of physics on laws is too narrow (see 1.3.3 and 1.3.4).

Introduction

The evidential gap for physical causal claims

A scientist, or someone in a day-to-day context, who wants to make or use a causal claim needs to provide evidence for this claim. In the previous chapter, the information about causal relations was provided by manuals which we could trust. However, in many contexts, we do not have such a specific causal information source. Yet ensuring that our causal information is *warranted,* is a prerequisite for using the information. So we often need to look for *evidence* to support our causal claims. The straightforward candidate source for evidence to support physical causal claims, would be the laws of physics. Yet according to Norton and Russell, the laws of physics do not contain causal information. One of their important arguments is that the laws of physics are time-symmetric:

> [T]he future 'determines' the past in exactly the same sense in which the past 'determines' the future. (Russell 1912, p.15)

Moreover, the laws of physics are symmetric in a more general sense: many are mathematical equivalencies. Causal relations, on the other hand, are frequently asymmetric in several ways: causes precede their effects, manipulating the cause changes the effect but not vice versa etc. Mathias Frisch has already shown that causal assumptions play a significant role in the reasoning of theoretical physicists (see 1.2.1).

But we also make them in day-to-day contexts. However, if the asymmetric physical causal knowledge does not come from the laws, where does it come from?

There is an evidential gap between many physical causal claims and the laws of physics. Recall the first complication I identified in the previous chapter: ascribing a causal relation to a phenomenon requires information about the context, since the law allows derivations in both directions. It supports both the claim that A causes B and its converse equally well. However, when we want to use causal information to intervene in phenomena (like in the previous chapter) or *explain* them (the topic of this chapter), we need to be certain of the causal information. In this chapter, I study how we support physical causal claims in the context of explanations.[2] My main question can be formulated as follows:

> How are the causal claims in physical causal explanations to be supported?

I will argue that it is mechanistic evidence, in combination with the laws of physics, that allows us to make substantiated causal claims which can be used for explanations. More specifically, this chapter will serve to do two things:

> (1) Showing that mechanistic evidence (information about the underlying mechanism) plays a crucial role in filling the evidential gap between physical laws and physical causal claims that we encounter when trying to build explanations.
>
> (2) Explicating how the two kinds of evidence interact with each other so that they can provide good reasons for accepting physical causal claims.

The reason I focus on physical causal claims that occur in explanations, is twofold. On the one hand, I have not discussed this type of using causal information yet. And second, this is a well-known topic in philosophy of the special sciences – especially in connection to evidence. So I have a lot of philosophical literature to guide me. At the same time, this will allow me to sketch the contrast between the discussion of evidence in the special sciences and the absence of this topic in philosophy of physics – which is less extreme than one may assume. This is related to the difference between regularities in the biomedical and social sciences on the one hand, and the laws of physics on the other. By reflecting on the different roles of mechanistic evidence, I also reflect on the differences

[2] I will only be discussing causal explanations. However, not all explanations are causal. I get back to this in 3.1.1.

between these regularities. Though the regularities from the special sciences are different from the laws of physics, I will show that this is more a matter of degree than a fundamental difference. This reflection will provide the setup for the next two chapters (chapter 4 and chapter 5) in which I study the relation between physical causal claims and the laws of physics more thoroughly.

Two types of evidence

Before I start my analysis, let me explain what I mean with 'correlational evidence' and 'mechanistic evidence'. By correlational evidence, I refer to information about the existence of a correlation in the world that can be used to argue for or against a causal claim. By mechanistic evidence, I refer to information about the existence of a mechanism in the world that can be used to argue for or against a causal claim. A *correlation* between variables **A** and **B** generally refers to a statistical connection between the two. This connection can be because of the following:

- A causes B
- B causes A
- Some (possibly unknown) factor C causes B and A
- Nonsense correlation

The first two options are rather obvious. For instance, if A causes B, whether this be in a deterministic or probabilistic way, the occurrence of A will raise the probability of B. Hence, there is a statistical connection. The third option expresses that both A and B are consequences of some common cause C. In such a case, it is the occurrence of C that raises the probability of both A and B. However, this also results in a statistical connection between A and B. I use the last option, a nonsense correlation, to collect all non-causal ways in which A and B can be connected. As for instance Jon Williamson (2005) discussed, this is in fact a large group:

> In fact probabilistic dependencies arise not only via causal connections, but also accidentally or because the variables are related through meaning, through logical connections, through mathematical connections, because they are related by (non-causal) physical laws, or because they are constrained by local laws or boundary conditions. (p.52)

All of these options, however, involve that there is no causal relation between A and B (or some common cause C). So from the perspective of supporting causal claims that we want to *use*, all of these options fail. I will therefore refer to all of them as "nonsense correlations". Note that from this quote, it is also clear that physical laws indeed carry

correlational evidence. This does not mean that the laws of physics are equal to correlations: they have many different properties. However, they are similar in at least one respect that is important for my point, viz. they express connections between variables. What I am mainly interested in here, is not how correlations and laws differ, but how mechanistic evidence interacts with laws. And this is analogous to how mechanistic evidence interacts with correlations. So I will treat laws as analogous to correlations in the sense that they both express connections between variables. In this way, laws can be seen as carrying correlational evidence.

Mechanistic evidence, on the other hand, refers to knowledge of a *mechanism*. The concept of a mechanism has become more and more important in philosophy of science, starting with Bechtel and Richardson's book on mechanisms in biomedical sciences in 1993 (Bechtel and Richardson 1993) and Elster's paper (1998) in the social sciences. Since then, the debate has kept on growing and the concept of mechanisms has been introduced to study almost all scientific disciplines. Like with almost all concepts I have discussed so far, many definitions have been suggested in the literature. Phyllis Illari and Jon Williamson have collected them and presented a pragmatic definition that can be used to study different disciplines:

> A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon. (2012, p.123)[3]

This is a definition that can be adapted to fit more specific scientific contexts when needed. This is the one I will be using (and where necessary specifying) throughout the dissertation (see also chapter 4).

Note that I distinguish between the evidence used to establish a correlation (like the results of randomised experiments, of prospective and retrospective studies, of correlational studies,…) and correlational evidence, in which the existence of a correlation is seen as established and used as evidence for or against a causal claim in an argument. In the same way, the evidence for or against the existence of a mechanism should be distinguished from mechanistic evidence as defined above. Illari discussed this difference aptly in (Illari 2011). In this chapter, I mainly focus on correlational and mechanistic evidence, but it goes without saying that the two other topics are extremely important and even a prerequisite for my analysis. Especially the evidence for a

[3] This definition closely resembles the "minimal mechanism" definition used in (*The Routledge Handbook of Mechanisms and Mechanical Philosophy* 2017), taken from (Glennan forthcoming). Because of the resemblance, I believe Glennan's definition could also be used here.

correlation will turn out to be crucial for deciding the informativeness of the correlational evidence.

3.1 Physical explanations

Let's come back to the well-known example of the flagpole that I mentioned in chapter 1 (see section 1.2.2). I suggested that

Exp_1: Explaining the length of the shadow by means of the height of the pole and the position of the sun is a good causal explanation.

Exp_2: Explaining the height of the pole by means of the length of the shadow and the position of the sun is a bad causal explanation.

Exp_3: Explaining the position of the sun by means of the height of the pole and the length of the shadow is a bad causal explanation.

I also suggested that these judgements are based on the following causal belief:

C_1: The position of the sun and height of the pole are causally relevant for the length of the shadow, but not the other way around.

Following my discussion in the previous chapter, it is not straightforward what this causal claim means. However, the weakest definition I gave can serve as a starting point here: both the position of the sun and height of the pole are positive causal factors for the length of the shadow. Because this is not a qualitative claim, but a quantitative one, my definition needs to be adapted. Recall D_1 from the previous chapter:

PCF: C (as opposed to C*) is a positive causal factor for E in the collection of setups U if and only if $\mathbf{P}_X(E)$ is greater than $\mathbf{P}_K(E)$.

To incorporate more quantitative claims, we can redefine C and E. They can refer to *changing* the value of the specific variables. In the case of C_1, this implies that C is defined as an increase in height of the pole (as opposed to keeping the height constant) and E is an increase in length of the shadow. So X is the hypothetical set of setups where the heights of the pole are increased, and K is the hypothetical set of setups where all the heights are kept constant. Finally, U is the collection of setups of type S_5:

(S_5) (1) The flagpole (with as variable feature its height), (2) the sun (with as variable feature its position), and (3) the shadow (with as variable feature its

length), with (1), (2) and (3) organised such that the flagpole is mounted in the ground and such that the sun shines on the flagpole.

From the definition of a mechanism, it should be clear that it bears some resemblance to the concept of physical setup. And indeed, the relevant information contained in the setup is often the same information that is needed to characterise the mechanism. In the context of my dissertation, the goal we have in describing physical setups is different from the goal we have in describing mechanisms. Describing the physical setups helps to make the domain of causal claims explicit. Describing a mechanism in the context of mechanistic evidence helps to determine whether a claim holds. Chapter 4 will show that the two are related, since it contains a mechanistic procedure to extrapolate causal knowledge. But for now, it suffices to see that mechanisms and physical setups perform different functions in my argument.

The judgement that C_1 holds is mainly based on intuition. How should we warrant that Exp_1 is a good explanation, and correspondingly, that causal claim C_1 can be trusted? How do we bridge the evidential gap? The first thing that needs to be settled, is what a good explanation consists of.

3.1.1 Pragmatic explanation

What is an explanation? Like with causation, there are too many definitions and accounts to count or discuss. For my pragmatic project that is focused on causation, I will be looking at physical causal explanations in a pragmatic way.[4]

The pragmatic toolbox for explanation developed by Erik Weber, Jeroen Van Bouwel and Leen De Vreese (2013) can be used to sketch how the explanations that I focus on can be understood. Following this framework, the explanations can be pragmatically characterised as answers to why-questions. In a pragmatic framework, the specificities of the why-questions also shape the answer. Weber, Van Bouwel and De Vreese distinguish between questions about facts and questions about regularities (2013, p.40). Here, I will only discuss questions about facts, but chapter 4 will show that these can be generalised to questions about regularities. Weber, Van Bouwel and De Vreese further distinguish three types of questions about facts: questions about plain facts, contrastive questions and resemblance questions (2013, pp.40-42). I will focus on questions about plain facts. In the case of the flagpole, this is the question

[4] This is not to say that all (physical) explanations need to be causal.

Why is does the length of the shadow (l) equal L?[5]

The way we answer this, is by combining (causal) information relevant for the case we want to explain and using this information to build an argument. As will become clear from the examples, this argument is often not deductively valid (see 3.1.3).

3.1.2 Modelling the phenomenon

The second point I want to reflect on is the way we select the boundaries and relevant parts of the phenomenon we want to explain. In 2.1, I already argued that the validity and meaning of the causal claims I am concerned with depend on the domain they refer to. I presented a way of making this domain explicit, viz. the concept of physical setups. However, except for a small note about relevant properties, I did not reflect on how we choose the parts of the setup. The way I proceeded was by choosing the setup in such a way that the causal claim underpinning remedy instructions would be valid. This strategy worked because I was looking for ways to make explicit the causal knowledge underlying efficacious remedy claims. However, when we want look for evidence for causal claims (like the ones functioning in explanations), we have no certainty about whether the causal claim holds.

So how do we select the objects that are part of the setup in the case of the flagpole, for instance? To characterise the interaction in the flagpole example, we need to see it as taking place in the setup composed of the sun, the flagpole and the ground. We abstract away from all the irrelevant factors and keep the relevant ones. At the same time, we abstract away from the irrelevant features of the elements in the setup. These decisions do not arise out of nowhere.

A debate that is relevant for this question deals with the way we *model* phenomena. The role that models play in science has a "rich and varied history", starting with people like James Clerk Maxwell and Lord Kelvin (Morrison and Morgan 1999a, p.1). An important contribution to the reflection on models is the book *Models as mediators*, edited by Margaret Morrison and Mary S. Morgan (1999c). They define models in the following way:

[5] As I showed in the previous chapter, there are often good reasons to introduce contrasts when discussing causal claims. For clarity, I will omit them here, but they can be added without much trouble.

> Models may be physical objects, mathematical structures, diagrams, computer programmes or whatever, but they all act as a form of instrument for investigating the world, our theories, or even other models. (Morrison and Morgan 1999b, p.32)

The main topic of the book is the relation between models, theory and data: models *mediate* between theory and data. The contributors argue that models should be seen as in some sense independent of the other two (Morrison 1999, p.43). The book addresses many important questions, but what I mainly want to focus on is the need to construct a model in order to reason about a phenomenon (viz. make causal claims, explain and intervene in it). The discussion on how models are built can shed light on how a physical setup is delineated.

Morrison and Morgan's book shows that building a model is a *practice fitting together relevant parts*. So firstly, it's a practice, the theory does not give you an algorithm to build a model (1999b, p.16). And second, the practice of modelling can be characterised as combining parts that we consider relevant for the task:

> […] [M]odels are built by a process of choosing and integrating a set of items which are considered relevant for particular tasks. (Morrison and Morgan 1999b, p.13)

In the pragmatic framework that I am working in, it is no surprise that one phenomenon can be modelled in different ways, depending on the goal we have. I will get back to this in chapter 5 (see 5.3.3). In the case of causal explanation, we include the elements (and relevant features) in the setup that we believe to be relevant for the explanation of the phenomenon. This also fits with Mieke Boon and Tarja Knuuttila's discussion of models in the engineering sciences. On their view, models are epistemic tools to reach specific goals (Boon and Knuuttila 2009, p.689). In a sense, models are reasoning aids.

What does all of this mean for the flagpole example? When we construct a physical setup, this is part of modelling the phenomenon and is based on what information we think is relevant for the explanation. Why are the sun and the flagpole included in the setup and not, say, some other object? Because we know that shadows are produced by light. So the choice to include the sun is based on background knowledge, more specifically, on judgements about which knowledge may be relevant for the explanation. I suggest that this is also why we include the information about the position of the sun: we know the relevant law of geometrical optics and believe that it might be useful for explaining the relevant phenomenon of the flagpole example. So we construct a physical setup in such a way that we believe we can explain the phenomenon. Correspondingly, during the process of explanation, we may change the setup to account for difficulties we encounter in our attempts to explain the phenomenon. Note that the modelling of the phenomenon we are explaining also includes decisions about the amount of detail we want to include. In chapter 5, I will pay more attention to how different demands

such as specificity or intelligibility can determine the regularities we use to explain phenomena.

What I mainly want to argue is that the picture from the previous chapter, where the setup was seemingly clear and the validity of the causal claim depended on the physical setup is more complicated. The way we conceptualise the phenomenon is not independent of the way we explain it and nor, as I will show, of the way we look for evidence for causal claims in the explanation. Now that this is straightened out, I can move on to how we look for evidence for causal claims in the explanations.

3.1.3 Explaining the flagpole

If explanations are answers to why-questions, the relevant why-question in the flagpole example is the following:

Why does the length of the shadow (l) equal L ?

The answer is given by the following explanation (Exp_4):

(C_9) In setups of type S_5, the height of the flagpole (h) and the position of the sun (α) determine the length (l) of the shadow according to $h/l = \tan\alpha$

The phenomenon we want to explain pertains to a setup of type S_5

$h = H, \alpha = \alpha_1$

*****

The length of the shadow $L = {}^{H}/_{\tan\alpha_1}$.

I use this starred line (*****) to separate the explanans from the explanandum because the explanation should not be seen as a deductively valid argument. In some cases, the explanation might have this strength, but since the causal relations need not be deterministic, it is often not the case. I will get back to this in 3.3.1.

This explanation consists of a causal claim (C_9) and information about the setup we are talking about. So how do we support the causal claim part of this explanation? As I said in the introduction, I conceptualise this as combining pieces of information. One piece of information is the law of geometrical optics that we thought was relevant for this phenomenon, viz.${}^{h}/_{l} = \tan\alpha$.However, recall from 1.2.2 that the law itself is not enough to warrant C_9, since the same law can also be used to warrant

(C_{10}) The length of the shadow and the position of the sun determine the height of the flagpole.

(C_{11}) The height of the flagpole and the length of the shadow determine the position of the sun.

Not only are all these claims compatible with the law, they are all equally well supported by it. Based on the law alone, we have no compelling reasons to prefer C_9 over the other causal claims. Yet as I argued above, if C_{10} or C_{11} holds instead of C_9, Exp_4 is not a good explanation. The laws of physics give us *correlational evidence* in the sense I defined in the introduction. However, we need some information to determine which causal claim holds in this case.

3.1.4 Explaining the pressure cooker

The flagpole is an artefact in the sense I characterised in chapter 1. However, this is not crucial for the explanation in section 3.1.3; a similar explanation can be given for the shadow of a tree, for example. In this chapter, I present a more paradigmatic example of an artefact, viz. a pressure cooker. This common household artefact can be characterised as a rigid closed container. Note that this is already a modelling step: it is characterised in such a way in light of the gas laws that specify volume-change as a relevant factor.

We want to answer for instance the following question:

> Why does the pressure (p) exerted on the walls of the cooker equal P?

To answer this question, we turn to the laws of physics. We characterised the pressure cooker as a rigid closed container because we believe we can explain it via the laws of thermodynamics. This is partly because modelling the cooker in terms of molecules is way too complicated. Frisch has argued for this: macro-phenomena cannot be modelled via micro-regularities (2014, p.38). I will get back to this in chapter 5 (see 5.3.2). In general, the next chapters will reflect more on the relation between artefact phenomena and (fundamental) laws of physics. For now, it suffices to treat the modelling of the pressure cooker in terms of laws of thermodynamics as a pragmatic choice.

When the cooker is heated, we believe the following causal claim holds:

> In pressure cookers, the temperature of the gas influences the pressure it exerts on the walls of the container.

How can we warrant this claim? We might try with the ideal gas law:

(IGL) For equal quantities of gas in a container, the product of the pressure p and the volume v is proportional to the temperature t, with a proportionality constant R (the ideal gas constant).

Or in its mathematical form:

$$pv = nRt$$

With p the pressure, v the volume, n the amount of substance of gas, R the ideal gas constant and t the temperature. However, to function as intended, a pressure cooker also needs to contain liquid. To describe a rigid container with liquid, a van der Waals equation is better:

$$\left(p + {an^2}/{v^2}\right)(v - nb) = nRt$$

This is a modification of the ideal gas law. The factor b expresses the volume per mole that is occupied by the molecules of the fluid. Factor a is a constant whose value depends on the gas. This law can be used to argue for the following causal claims – as correlational evidence[6]:

(C_{12}) For pressure cookers, the temperature of the gas combined with the volume of the gas determines the pressure it exerts on the walls of the container.

(C_{13}) For pressure cookers, the pressure exerted on the walls of the container combined with the volume of the gas determines the temperature of the gas.

(C_{14}) **F**or pressure cookers, the temperature of the gas combined with the pressure exerted on the walls of the container determines the volume of the gas.

We do not have enough information to decide what claims to accept. Yet to explain the amount of pressure the gas exerts on the walls of the cooker, we need C_{12} to hold:

> In pressure cookers, the temperature of the gas (t) influences the pressure (p) it exerts on the walls of the container according to $\left(p + {an^2}/{v^2}\right)(v - nb) = nRt$
>
> $t = T, v = V$
>
> *****
>
> The pressure exerted on the walls of the cooker is $P = \frac{nRT}{(V-nb)} - \frac{an^2}{v^2}$

Otherwise, the explanation does not work. How do we bridge this evidential gap? What type of information do we need to add, so that we can present an argument that allows us to accept C_{12}? To answer this question, I will take inspiration from the way similar

[6] It is important to note that each of C_{12} till C_{14} is equally supported by the law. This again shows the tension between symmetric laws and asymmetric causal claims.

problems are handled in philosophy of the biomedical and the social sciences. In these domains, knowledge of *mechanisms* is invoked in combination with probabilistic evidence to argue in favour of a causal relation (see also 3.4.3). The literature on mechanistic evidence for causal claims in the biomedical sciences is quite extensive. I will use this to sketch the main strategy. In section 3.3, I will then explain a similar way this works in the social sciences. However, the two are not completely the same. By reflecting on the difference, I will be able to explain why the social sciences are a better guide for understanding the physical causal claims. Finally, in section 3.4, I will come back to the physical examples and reflect on the differences between the domains.

3.2 Mechanistic evidence in the biomedical sciences

Evidence for causal claims in biomedical sciences, especially the interplay between mechanistic and correlational evidence, has received a lot of attention from philosophers in the past years. I present an example of the biomedical sciences where correlational and mechanistic evidence are both used. I then discuss some of the philosophical literature on this topic to analyse the evidential situation in the biomedical sciences.

3.2.1 A biomedical example

Suppose that John has heart disease and we want to know why. In the biomedical sciences, the following explanation could be given:

> One of John's legs was amputated.
>
> Limb amputation is a positive causal factor for heart disease.
>
> *****
>
> John has heart disease

Contrary to physics, biomedical science explicitly looks for causal regularities:

> Biomedical science is a term used to describe the study of the causes, consequences, diagnosis, and treatment of human diseases. (Ahmed, Glencross, and Wang 2011, p.19)

However, establishing a causal claim in the biomedical sciences is not that easy. Many studies have been done in order to establish a causal connection between leg amputation and cardiac arrest or heart disease. See for example (Hrubec and Ryder 1980) and (Modan et al. 1998). The way these studies go, is first by performing a *prospective cohort study*. In a cohort study, a group of individuals with a particular characteristic is followed over a period of time (Illari and Russo 2014, p.11). In this case, the 'characteristic' is that the participants' limb was amputated. The researchers followed the participants in order to determine whether they suffer on average more from cardiovascular disease than people without amputated limbs. If this is the case, then a *correlation* is established. However, like I explained in the introduction, variables can be correlated for a number of reasons. To establish a causal claim, correlational evidence is often not enough.

In the Hrubec and Ryder article, the authors explicitly say that they do not know the reason for the "statistically relevant relationship" between limb amputation and cardiovascular disorders (1980, p.247). Modan et al. also found a statistically relevant relation between limb amputation and cardiovascular disorders. They suggest several possible mechanisms for this and other related connections:

> To link these observations, it is conceivable that in amputees chronic mental and physical stress, with their attendant increased autonomic nervous system activity, could augment blood coagulability, either directly or indirectly, through hemodynamically enhanced shear stress forces. The latter play a role in the initial injury to the vascular endothelium preceding atheroma formation as well as in activation of plate sensitivity and blood viscosity, across a range of shear rates. (Modan et al. 1998, pp.1244-1245)

Recall Illari and Williamson's pragmatic definition of a mechanism that I quoted in the introduction. The description above is that of a mechanism: it contains entities (such as autonomic nervous system, blood, vascular endothelium) with activities (augmenting blood coagulability, coagulating, breaking down) in a specific organisation (blood in the veins, vascular endothelium lines the inside of blood vessels). With this description, the authors attempt to account for the statistical connection. Modan et al.'s article is not an exception in the biomedical sciences: mechanistic evidence is often invoked to account for the statistical results of experimental or observational studies. This has not gone unnoticed by philosophers of science, as I will show in the next section.

3.2.2 Mechanistic evidence in the biomedical sciences

As I explained in the introduction, the concept of mechanism has been on a rise in philosophy of science since the nineties. Mechanistic evidence however, constitutes a topic on its own. In 2007, Russo and Williamson published a now famous paper on evidence in the biomedical sciences. The central thesis, or at least the most cited one, was dubbed the Russo-Williamson thesis and captured by this quote:

> To establish causal claims, scientists need the mutual support of mechanisms and dependencies. The idea is that probabilistic evidence needs to be accounted for by an underlying mechanism before the causal claim can be established. (Russo and Williamson 2007, p.159)

So Russo and Williamson clearly plead for a significant role for mechanistic evidence in the biomedical sciences. The specificities and modalities of this plea, however, have been highly debated. Philosophers have interpreted the thesis in several ways and have thoroughly examined all of the interpretations (see for instance (Weber 2009), (Leuridan and Weber 2011), (Gillies 2011), (Broadbent 2011)). As Illari argues, some formulations in the paper did allow some ambiguous interpretations (2011). However, Russo and Williamson have, in a later article, clarified what they mean:

> According to the epistemic theory, causal claims need to be made on the basis of evidence of both difference-making and mechanisms, as well as evidence such as temporal information and information about the nature of the events in question (Russo and Williamson 2011, p.568)

This is still a pretty strong claim, since both types of evidence are required to argue for causal claims. I do not commit to such a strong claim. What I want to take from this debate, is that in philosophy of the biomedical sciences, it is accepted that mechanistic information is often invoked in favour of a causal claim. There may exist disagreement on the *necessity* of this mechanistic evidence, but philosophers generally agree that mechanistic evidence is useful to support causal claims in the biomedical sciences. In the next section, I focus on what this mechanistic evidence specifically does.

3.2.3 Which evidential gap?

Like with the physical causal claims, the biomedical scientists start from correlational evidence. Most of the time, this is obtained from observational studies or experimental studies. Observational studies refers to methods of data-gathering where we passively record observations (Illari and Russo 2014, p.10). Examples of observational studies are

cohort-studies (like the example above) and case-control studies (where two individuals are compared) (Illari and Russo 2014, p.11). In experimental studies, on the other hand, data are produced "within the experimental setting by manipulating some factors while holding others fixed" (Illari and Russo 2014, p.10). RCT's (Randomized Controlled Trials) are a well-known example of experimental studies.

What is inherently present in the data gathered by those studies, is the *temporal direction.* In prospective studies, we follow people with a certain condition to see whether they develop a certain other (previously absent) condition. This often allows us to conclude that the latter is not the cause of the first. In experimental studies, one variable is actively manipulated and the value of the other variable is measured. Here too, the temporal direction is clear. Because we believe that causes precede their effects, there is no question of which variable caused the other. In the introduction, I mentioned four possible reasons for a correlation between A and B: a causal direction from A to B, a causal relation from B to A, some unknown common cause for both A and B, and a nonsense correlation. In relation to these four options, in the biomedical sciences, we can often exclude one of the causal relations because of temporal considerations regarding how the correlational evidence was gathered. So the mechanistic evidence is, in this context[7], often invoked to decide between the remaining causal direction and a common cause or nonsense correlation.

So the biomedical sciences is not the best role model to understand the role of mechanistic evidence for physical causal claims, since the distinction between two possible causal directions (A causes B or B causes A) was very central for the latter (I will reflect on the possibility of common causes and nonsense connections in 3.4.3). In the next section, I will reflect on mechanistic evidence in a case from the social sciences. The role that mechanistic evidence plays there, is more related to what we need for the physical causal claims. This case will allow me to pinpoint what this role consists of.

[7] It's important to note that mechanistic evidence can also be used for other goals, like extrapolation (see chapter 4).

3.3 Mechanistic evidence in the social sciences

For the social sciences, I will look at Duverger's laws. I first explain what these laws are and then analyse how Duverger argued for them. I will argue that this too is an example of mechanistic reasoning. Moreover, this reasoning can help us understand what mechanistic evidence can do for physical causal claims, since the direction of causation is a problem here as well.

3.3.1 Background

The French political scientist Maurice Duverger became famous in the 1950s for his work on the relation between electoral systems and the number of political parties. The key propositions are the following ((Benoit 2006, p.70), (Duverger 1959, pp.217 and 239)):

(P_1) The simple-majority single-ballot system favours the two-party system.
(P_2) The majority system with a second-round runoff favours multi-partism.
(P_3) Proportional representation favours multi-partism.

In a simple-majority single-ballot system there is one member of parliament to be elected in each voting district. The candidate who gets more votes than any other candidate is elected (even if there is no majority, i.e. even if the winner's score is less than 50%). Duverger considers two other systems: the majority system with a second-round runoff (if no candidate receives more than 50% of the initial votes, there is a second round with the top-two candidates) and proportional representation (multiple members of parliament for each district; seats allocated based on percentage of votes for each political party).

How should we understand the causal claims? I argued that the physical causal claims I am concerned with are implicitly or explicitly about a domain. The same holds for Duverger's claims. In the case of Duverger, the domain that he and his fellow political scientists are talking about is the set of all democratic countries.

To make the meaning of the causal claim explicit, the definitions from the previous chapter come in handy. Recall the idea of hypothetical populations used in the definition of a positive causal factor (D_1). To recapitulate, C as opposed to C* is a positive causal factor for E if, in the hypothetical population obtained by changing all C* into C, we have more E than in the hypothetical population where we changed all C to C*. Taking this into account, the meaning of proposition P_3 can be explicated as follows:

If all democratic countries were to have proportional representation, there would be more countries with a multi-party system than if all democratic countries were to have a simple-majority single-ballot system.

This is a non-deterministic or probabilistic causal relation. Not every democratic country with a simple-majority single-ballot system has a two party system (Canada and India are exceptions); nor is it the case that all countries with proportional representation have a multi-party system (Austria is an exception). This is not a problem, many causal claims throughout the sciences are probabilistic, like the causal relation between smoking and lung cancer. Wanting to base all causal explanations on deterministic causal relations is not realistic. Recall from the first chapter that finding sufficient causes is not an easy task even in artefacts that we designed. In social and biomedical sciences, it is even more difficult, since we have less control. So explanations are often based on probabilistic causal claims. This is also why, as I said in 3.1.3, the explanation is not necessarily a deductively valid argument.

Now that this is settled, let's turn to the explanations. Belgium has a multi-party system. To explain this, we need to answer the question:

Why does Belgium have a multi-party system?

By means of P_3, we can answer this question as follows:

Belgium has a proportional representation system.

Proportional representation favours multi-partism.

*****

Belgium has a multi-party system.

How did Duverger argue for P_3?

3.3.2 Correlational and mechanistic evidence

Duverger performed an extensive *comparative study* of the relation between electoral systems and number of parties. A comparative study consists of comparing certain properties across different countries or cultures. Duverger's comparative study provided evidence for several correlation claims, like:

(P_4) Proportional representation is positively correlated with multi-partism.

(P_5) The simple-majority single-ballot is positively correlated with a two-party system.

The results of a comparative study do not fix the temporal order in the same way as certain ways of collecting correlational evidence in the biomedical sciences did. The correlational data is less informative. In order to be able to give causal explanations, we need evidence that suffices to accept the causal claims P_1 – P_3. The correlational evidence is not enough, we need to exclude the three alternative options: the reverse causal direction, an unknown common cause, and nonsense correlations.

The way Duverger argues for a specific causal direction is by invoking what he calls 'the mechanical effect' and 'the psychological factor':

> The mechanical effect of electoral systems describes how the electoral rules constrain the manner in which votes are converted into seats, while the psychological factor deals with the shaping of voter (and party) responses in anticipation of the electoral law's mechanical constraints. (Benoit 2006, p.72)

The mechanisms Duverger points at are *social mechanisms.* Steel characterises these as follows:

> Social mechanisms are complexes of interacting individuals, usually classified into specific social categories, that generate causal relationships between aggregate-level variables. (2004, p.59)

Steel's definition can be seen as a specification of Illari and Williamson's definition (see the introduction), tailored for the social sciences.

3.3.3 Which evidential gap?

Let me look at the first mechanism in more detail. When polling stations are closed, votes are counted. This is done by people (with all kinds of technological assistance) who perform certain roles in the electoral system. These roles are the "social categories" which Steel refers to, for instance "chairman of totalisation office" or "secretary" or "assessor" in such offices. The interaction between all these people who count and process votes in a predetermined, highly structured way leads to the proclamation of a result (again by an individual with a specific social role) in terms of seats in the parliament.

This whole process can be seen as an input-output-system in which votes are processed according to certain electoral rules and result in a distribution of seats. There is always a certain mismatch between share of votes and share of seats:

> The mechanical effect of electoral systems operates on parties through the direct application of electoral rules to convert votes into seats. In the mapping of vote

> shares to seat shares, some parties — almost always the largest ones — will be 'over-represented,' receiving a greater proportion of seats than votes. Because this mapping is a zero-sum process, over-representation of large parties must create 'under-representation' of the smaller parties. (Benoit 2006, p.73)

However, in simple majority single-ballot electoral systems, application of the electoral rules leads, on average to higher over-representation of large parties.

An important aspect of this mechanism is that it is causally directed. The people involved do not count seats and convert them into votes. They count votes and convert them into seats. And electoral rules are also an input of the process, not an output. The rules are in the heads of the individual (and implemented in the computer programmes they use). They are in no way an output of the vote processing system.

About the second mechanism, Benoit writes:

> Duverger's psychological effect comes from the reactions of political actors to the expected consequences of the operation of electoral rules. The psychological effect is driven by the anticipations, both by elites and voters, of the workings of the mechanical factor, anticipations which then shape both groups' consequent behavior (Blais and Carty, 1991, 92). Under electoral rule arrangements that give small or even third-place parties little chance of winning seats, voters will eschew supporting these parties for fear of wasting their votes on sure losers. Political elites and party leaders will also recognize the futility of competing under certain arrangements, and will hence be deterred from entry, or motivated to form coalitions with more viable prospects. (2006, p.74)

This represents another social mechanism in the sense of Steel: there are individuals with certain roles ("party leader", "voter") behaving and interacting in certain ways. The mechanism rests on two assumptions about how relevant behaviour is determined:

> - What voters do in the polling station is influenced by their knowledge of the electoral system and the degree to which it favours large parties.
>
> - What party leaders do in terms of pre-election coalitions is influenced by their knowledge of the electoral system and the degree to which it favours large parties.

The mechanistic information brings in a specific temporal and with that causal order: the electoral rules exist, they influence what is in the mind of voters and party leaders. And the opinions of voters and party leaders determine their behaviour. So in the case of Duverger, the temporal inference is quite convincing. Note that comparative studies can also be used in the biomedical sciences, and social scientists often perform prospective and retrospective studies as well. The role that mechanistic evidence plays, is less related

to the scientific disciplines than to the ways the correlational evidence was gathered. However, in the biomedical sciences, it is more likely that the way we collected the correlational evidence allows us to exclude a possible causal direction. Cases like the one described by Duverger are more common in the social sciences.

The mechanistic evidence is also used to strengthen our belief in the causal relation, like it does in the biomedical sciences (see 3.2.2). But more importantly for my purposes, it also helped decide the *direction* of the causal relation. In the next section, I will show that mechanistic evidence can do the same for physical causal claims and in this way, bridge the evidential gap.

3.4 Mechanistic evidence for physical causal claims

I can now revisit the physical examples from 3.1: the flagpole and the pressure cooker. In 3.1, I explained that we model the phenomenon in such a way that it can be captured by a law of which we believe it can help to explain the phenomenon. Most of the physical laws, however, were not enough to argue for the causal claims on which the explanations were based. They only provide what I called correlational evidence: they tell us of a connection between A and B, but are not informative enough to exclude all the possible alternatives to one specific causal relation (see the introduction). Mechanistic evidence can be invoked to choose the right alternative, like it was in Duverger's case.

3.4.1 Revisiting the flagpole

For the flagpole we needed to decide between the following causal claims:

(C_9) The height of the flagpole and the position of the sun determine the length of the shadow.

(C_{10}) The length of the shadow and the position of the sun determine the height of the flagpole.

(C_{11}) The height of the flagpole and the length of the shadow determine the position of the sun.

Information regarding how the shadow comes to be in that system, can help us decide:

> The sun emits light rays. They travel to earth and are blocked by the flagpole. This results in the dark area called the shadow.

To see that this is mechanistic evidence, recall the pragmatic definition of a mechanism by Illari and Williamson that I mentioned in the introduction:

> A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon. (2012, p.123)

In this example, the entities are the sun, the pole and the ground. The activities are emitting light (by the sun) and blocking light (by the pole). They are organised such that the sun makes an angle α with the ground. Information about these entities and activities that are responsible for the shadow in the system described S_3, combined with the law of geometrical optics, does give us a compelling reason to accept C_9 instead of C_{10} or C_{11}. If the mechanism were different, we could accept a different causal claim, and correspondingly, give a different explanation for the phenomenon.

This bears a lot of resemblance to the role mechanistic evidence played in the case of Duverger: the mechanism is causally directed and allows us to decide which variable influences the other.

3.4.2 Revisiting the pressure cooker

In the case of the pressure cooker, we needed a way to argue for C_{12}, instead of C_{13} and C_{14}.

(C_{12}) The temperature of the gas combined with the volume of the gas determines the pressure it exerts on the walls of the container.

(C_{13}) The pressure exerted on the walls of the container combined with the volume of the gas determines the temperature of the gas.

(C_{14}) The temperature of the gas combined with the pressure exerted on the walls of the container determines the volume of the gas.

Mechanistic evidence can help here as well. Specifically, we can use the following information:

> We fill and seal the pressure cooker. We heat the pressure cooker by means of a stovetop. This heating results in an increase in temperature of the liquid. The water comes to boil and turns into gas, until the water and gas are in equilibrium. The heat increases the energy of the gas, resulting in a rise in pressure the gas exerts on the walls of the container.

This is again a mechanism: the entities are the stovetop, the liquid, the gas and the container. The activities are for instance heating, turning into gas, increasing in energy. The organisation is such that the liquid is in the cooker, the cooker is on the stovetop, etc. The pressure cooker is a typical example of a man-made physical setup. For artefacts that function properly, it is easier to decide how the mechanism works than in the case of natural examples, since we have more control over artefacts: we designed them (see also 4.1.2).

So mechanistic information is needed to determine the causal direction in specific setups and correspondingly, to help us decide which of the causal claims that are compatible with the law, to accept.

3.4.3 The different evidential gaps

I can now reflect on the difference between the role of mechanistic evidence for the physical claims compared to the social and biomedical claims. This is related to what I mentioned about the laws being analogous to correlations, but not equal. The main purpose of the mechanistic evidence in case of the physical causal claims, was to decide whether A caused B rather than B caused A. This is significantly different from the role of mechanistic evidence in the other examples. That is because of the nature of the laws of physics: many are mathematical equivalencies. As such, the laws allow us to determine the value of every variable based on the value of the other variables in the law.[8] Moreover, I will show that this also has the consequence that we can often exclude the possibility of unknown common causes and nonsense correlations. Let me begin with the latter. For the laws I discussed (paradigmatic laws of physics, that are thought to hold for the rather idealised systems I discuss), if we believe that the physical law holds, then the option of a nonsense correlation is excluded from the start. A physical regularity is given the status of *law* exactly because the connection between the variables it expresses, is sensible.

Moreover, for many physical laws unknown common causes are also excluded. If you believe that a physical regularity expressed as an equivalency is applicable to the system and phenomenon in question[9], for many laws, we can exclude common causes that are not variables in that regularity. Many laws attempt to include all factors that influence

[8] Some laws also contain constants. We can also calculate the value of those based on the values of the other members of the equations.

[9] Recall that we modelled the phenomenon in such a way that the law would be applicable. See 3.1.2.

the outcome, so we can often use this information to refute the possibility of unknown common causes. If a factor that influenced the outcome was not included in the regularity, there would be no equivalency. So if you believe a physical law expressed as an equivalency is applicable to your setup, this law also determines which variables can potentially occur as cause variables in true causal claims. The variables that mattered are all given by the law and the phenomenon was modelled in such a way that it could be captured by the law.[10] Other physical laws are probabilistic. In this way, they take disturbing factors into account. So they do not exclude that other factors can influence the result. On the contrary, they take them into account by only expressing a probability. Yet there is some implicit distinction between proper causes and disturbing background or shock factors. The proper causes are mentioned in the law, the disturbing factors are omitted. This marks a difference with the probabilistic regularities from the special sciences.

However, because the mechanistic evidence that is needed for the physical causal claims requires information about the setup, this mechanistic evidence only allows us to argue for a very *local* causal claim. Physical setups are highly specific. Compared to Duverger's claims (whose domain was all democratic countries), or the biomedical example from 3.1.2 (that held for all humans), the physical setups are less widespread, and there are more alternatives to the setups. So mechanistic evidence is used to argue for physical causal claims as well, but the claims that are argued for are significantly less general than the ones in the biomedical or social sciences. The claims are less stable across varying background conditions. I will get back to this in chapter 5.

Regardless of these differences, there are a lot of similarities between the examples from the special sciences and my physical cases. However, in the philosophy of the special sciences, invoking mechanistic evidence to support causal claims is significantly more accepted than it is in the philosophy of physics. Reasoning in terms of mechanistic evidence is not encountered a lot. Instead, problems are expressed in terms of mathematics, like initial-value problems. Qualitative reasoning does not receive a lot of attention. I hope my chapter has shown that there is something to be gained from reasoning in terms of mechanistic evidence for physical cases as well.

[10] This of course presupposes that there is enough evidence for the law itself. See also the introduction.

Conclusion

In this chapter, I discussed how we argue for physical causal claims in the context of physical explanations. Specifically, I argued that the laws of physics by themselves do not suffice to argue for physical causal claims. Because many are mathematical equivalencies, they can be used to determine the value of each variable based on the value of the other variables in the equivalency. They are symmetric. However, I showed that mechanistic evidence, in combination with the laws, does suffice to argue for physical causal claims. In this way, this chapter constituted an argument for my second specific claim (II). I used two examples: a flagpole and a pressure cooker. For the flagpole, we explain the length of the shadow as follows:

(C_9) In setups of type S_5, the height of the flagpole (h) and the position of the sun (α) determine the length (l) of the shadow according to $h/l = \tan \alpha$

The phenomenon we want to explain pertains to a setup of type S_9

$h = H, \alpha = \alpha_1$

*****

The length of the shadow $L = {}^{H}/_{\tan \alpha_1}$.

To argue for the causal claim in the explanation, I argued we can rely on a combination of (1) correlational evidence captured by the relevant law of geometrical optics, viz. $h/l=\tan \alpha_1$ and (2) mechanistic evidence provided by information about how the shadow comes to be, viz.:

> The sun emits light rays. They travel to earth and are blocked by the flagpole. This results in the dark area called the shadow.

In the pressure cooker example, on the other hand, we were interested in the value of the pressure. We explained it as follows:

> In pressure cookers, the temperature of the gas (t) influences the pressure (p) it exerts on the walls of the container according to $\left(p + {}^{an^2}/_{v^2}\right)(v - nb) = nRt$
>
> $t = T, v = V$
>
> *****

The pressure exerted on the walls of the cooker is $P = \frac{nRT}{(V-nb)} - \frac{an^2}{v^2}$

In combination with the Van der Waals equation mentioned in the causal claim, I argued that we needed the following mechanistic information:

> We fill and seal the pressure cooker. We heat the pressure cooker by means of a stovetop. This heating results in an increase in temperature of the liquid. The water comes to boil and turns into gas, until the water and gas are in equilibrium. The heat increases the energy of the gas, resulting in a rise in pressure the gas exerts on the walls of the container.

To argue that mechanistic evidence can fill the evidential gap for physical causal claims, I took inspiration from the philosophy of the biomedical and of the social sciences. Philosophers in the biomedical sciences have stressed the importance of mechanistic evidence to increase confidence in a causal relation. However, as I showed, in the biomedical sciences the mechanistic evidence is mostly relied on to exclude potential common causes and nonsense correlations. This is because most observational and experimental studies used in the biomedical sciences already give you temporal information. In the political sciences, on the other hand, comparative studies are omnipresent and the results of such studies do not contain temporal information. There, scientists use mechanistic evidence to argue for one specific causal claim instead of another (while also attempting to exclude common causes and nonsense correlations). But regardless of these differences, these literatures helped me show that mechanistic evidence brings the causal asymmetry to correlational evidence that is needed to warrant physical causal claims.

Mathias Frisch has made a similar point with regard to the reasoning of theoretical physicists. He argued that causal assumptions help to draw conclusions when dynamical models are underdetermined. Among other things, they bring in temporal asymmetry (Frisch 2014, p.127). He defined these causal assumptions in terms of Pearl's structural account of causation (Frisch 2014, p.235). While I pay particular attention to artefacts, Frisch focuses on theoretical physics. Nevertheless Frisch's and my projects line up: we both want to pay genuine philosophical attention to causal claims and reasoning about physical systems. And more specifically, we both show that temporal asymmetry can be brought in by means of certain assumptions regarding causal structures.

In general, evidence for physical causal claims is not often discussed in philosophy of physics. However, I have shown that finding evidence for causal claims is not easy. For one, the laws of physics do not suffice. So all the reflection on the laws of physics by philosophers of physics does not really help us in trying to understand how we provide evidence for physical causal claims. At the same time, my analysis in this chapter showed

that there are interesting philosophical issues related to evidence that arise when we look at using physical causal knowledge. In this way, it contributed to establishing my two generic aims, viz. (A) and (B).

The interesting issue I focused on, was what I called the evidential gap. But there are other aspects that make it even more difficult to find evidence for causal claims, aspects that I have only briefly reflected upon and taken for granted. One such aspect is the way we model a phenomenon. I discussed this in 3.1.2, by arguing that we model a phenomenon in a specific way so that it can be covered by a law that we think is relevant for the phenomenon. For the flagpole and the pressure cooker, this might look straightforward. However, in new phenomena, this is way less clear. If a phenomenon occurs for the first time, or is studied for the first time, or is studied in a more detailed way or with a different goal (recall that this can influence the meaning of the causal relation), it is not at all clear which laws might be relevant for explaining them. Yet this is important for modelling. Relatedly, in such contexts, getting the mechanistic evidence, is also not that easy. This is widely acknowledged in the philosophy of the biomedical sciences. In the case of artefacts like the pressure cooker, we know the mechanism, since we built it. This gives us more knowledge of the mechanism than in biomedical cases. However, when artefacts malfunction, we need different knowledge, since the functioning of the artefact is disturbed. As such, discovering which mechanism is active becomes more daunting. A final complicating factor is areas where the theory of physics is not as established as thermodynamics or geometrical optics. You might think of quantum physics, but one need not go as far. There are many domains in the engineering sciences that do not have laws as established as classical physics.

However, engineers use physical causal knowledge on a daily basis. So clearly, they succeed in modelling phenomena in such a way that they can make relevant causal claims about them, argue for these claims and use the causal information in a goal-directed way. How do they handle all these difficulties? In the next chapter, I will look at failure analysis, a specialisation in the engineering sciences, to investigate how less straightforward causal claims are argued for. Specifically, I will study the way they extrapolate physical causal knowledge from one context to another and the evidence they need for those extrapolations. This will help to understand how more complex physical causal claims are argued for, and simultaneously, further scrutinise the assumed central position of laws in philosophy of physics. As I announced, the following chapter will generate even more interesting philosophical issues. In this way, it will also contribute to further realising my two generic aims.

References

Ahmed, Nessar, Hedley Glencross, and Qiuyu Wang. 2011. *Biomedical science practice: experimental and professional skills*, *Fundamentals of biomedical science*. Oxford: Oxford University Press.

Bechtel, William, and Robert C. Richardson. 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton: PUP.

Benoit, Kenneth. 2006. Duverger's law and the study of electoral systems. *French Politics* 4 (1):69-83.

Boon, Mieke, and Tarja Knuuttila. 2009. Models as epistemic tools in engineering sciences: a pragmatic approach. *International Journal of Software Engineering and Knowledge Engineering*:687-720.

Broadbent, Alex. 2011. Inferring causation in epidemiology: mechanisms, black boxes, and contrasts. In *Causality in the Sciences*, edited by P. Illari, F. Russo and J. Williamson: Oxford University Press.

Duverger, Maurice. 1959. *Political parties: Their organization and activity in the modern state*. 2nd English Revised edn ed. London: Methuen.

Elster, Jon. 1998. A plea for mechanisms. *Social mechanisms: An analytical approach to social theory* 49.

Frisch, Mathias. 2014. *Causal reasoning in physics*: Cambridge University Press.

Gillies, Donald. 2011. The Russo-Williamson thesis and the question of whether smoking causes heart disease. In *Causality in the Sciences*, edited by P. Illari, F. Russo and J. Williamson: Oxford University Press.

Glennan, Stuart. forthcoming. *The New Mechanical Philosophy*. Oxford: Oxford University Press.

Hrubec, Zdenek, and Richard A. Ryder. 1980. Traumatic limb amputations and subsequent mortality from cardiovascular disease and other causes. *Journal of chronic diseases* 33 (4):239-250.

Illari, Phyllis McKay. 2011. Mechanistic Evidence: Disambiguating the Russo–Williamson Thesis. *International Studies in the Philosophy of Science* 25 (2):139-157.

Illari, Phyllis McKay, and Jon Williamson. 2012. What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science* 2 (1):119-135.

Illari, Phyllis, and Federica Russo. 2014. *Causality: Philosophical Theory meets Scientific Practice*. 1 edition ed: Oxford University Press.

Leuridan, Bert, and Erik Weber. 2011. The IARC and Mechanistic Evidence. In *Causality in the Sciences*, edited by P. Illari, F. Russo and J. Williamson: Oxford University Press.

Modan, Michaela, Einat Peles, Hillel Halkin, Hedva Nitzan, Morris Azaria, Sanford Gitel, Dan Dolfin, and Baruch Modan. 1998. Increased cardiovascular disease mortality rates in traumatic lower limb amputees. *The American journal of cardiology* 82 (10):1242-1247.

Morrison, Margaret. 1999. Models as autonomous agents. In *Models as Mediators*, edited by M. Morrison and M. Morgan: Cambridge University Press.

Morrison, Margaret, and Mary S. Morgan. 1999a. Introduction. In *Models as Mediators*.

———. 1999b. Models as mediating instruments. In *Models as Mediators*, edited by M. Morrison and M. Morgan: Cambridge University Press.
———. 1999c. *Models as Mediators*: Cambridge University Press.
The Routledge Handbook of Mechanisms and Mechanical Philosophy. 2017. Edited by S. Glennan and P. Illari. London: Taylor & Francis.
Russell, Bertrand. 1912. On the notion of cause. *Proceedings of the Aristotelian society* 13:1-26.
Russo, Federica, and Jon Williamson. 2007. Interpreting causality in the health sciences. *International studies in the philosophy of science* 21 (2):157-170.
———. 2011. Epistemic Causality and Evidence-Based Medicine. *History and Philosophy of the Life Sciences* 33 (4):563-581.
Steel, Daniel. 2004. Social mechanisms and causal inference. *Philosophy of the social sciences* 34 (1):55-78.
Weber, Erik. 2009. How Probabilistic Causation Can Account for the Use of Mechanistic Evidence. *International Studies in the Philosophy of Science* 23 (3):277-295.
Weber, Erik, Jeroen Van Bouwel, and Leen De Vreese. 2013. *Scientific Explanation*: Springer Science & Business Media.
Williamson, Jon. 2005. *Bayesian Nets and Causality: Philosophical and Computational Foundations*: Oxford University Press.

Chapter 4 From one to many: generalisation and evidence in failure analysis

In this chapter, I turn my attention to how engineers generalise physical causal knowledge in a way that they can use it. This requires them to provide evidence[1] for generalising the claims, and for determining the domain of their generalised claims. I will investigate how this evidence can be characterised. This is connected to claim (III).

Many scientific disciplines start from causal knowledge about a limited number of cases and attempt to draw conclusions about different or more cases. Recall Cartwright and Hardie's example of the nutritional project for pregnant women in Bangladesh that failed in Tamil Nadu that I discussed in section 1.4. This is a case from the social sciences. A similar method is used in biomedical sciences. In the context of these sciences, philosophers have attempted to understand how and when (causal) knowledge of one case or population can be *extrapolated* to other cases or populations. In this chapter, I will analyse cases from engineering science (specifically failure analysis) to argue that a similar problem holds for physical causal knowledge. However, as I will show, the problem is not entirely the same. Because artefacts are frequently analysed from what I call a *design-perspective*, the generalisations that take place are different than those in social sciences or biomedical sciences. I will therefore adapt philosophical tools given by Daniel Steel and Nancy Cartwright, to gain more insight in the inferences leading to these generalisations. This will result in a two-fold mechanism based heuristic for such inferences. Correspondingly, my account provides insight in the evidence needed for those inferences.

[1] Like in the previous chapter, I use evidence to refer to information that can be used to argue for (the domain of) causal claims. In this chapter, different kinds of scientific data are considered, such as measurements.

The causal knowledge that is produced in the cases that I study in this chapter differs in two respects from the knowledge I discussed in chapter 3. For one, in the current chapter, general causal knowledge is produced, while the cases in chapter 3 looked for particular causal knowledge. Second, this general knowledge is based in local knowledge, instead of laws, as was the case in chapter 3. What this chapter has in common with chapter 3 is its focus on evidence. As announced, the analysis of these more complex cases will introduce another set of complexities related to how physical causal knowledge is found and used in actual scientific practice. The generalisation of causal knowledge in light of applications involves a lot of reasoning and evidence – and the same holds for determining the domain of this generalised knowledge. By reflecting on how analysts succeed in this, the challenges become clear. Hence, this chapter will address yet another interesting philosophical issue related to using physical causal knowledge. In this way, my argument for (A) and (B) is gradually developing.

Introduction

When an artefact breaks down[2], specialised engineers called failure analysts study the specific circumstances that led to this failure. For instance, in "Creep failure of a spray drier", Paul Carter investigates the collapse of a specific spray drier[3] which had been in service for nearly 20 years (2001, p.73). This article was reprinted in *Failure Analysis Case Studies II*, a collection of "40 case studies describing the analysis of real engineering failures which have been selected from volumes 4, 5 and 6 of Engineering Failure Analysis" (*Failure Analysis Case Studies II* 2001, p.v).[4] In the preface, the editor comments on the previous edition:

> The book has proved to be a sought-after and widely used source of reference material to help people avoid or analyse engineering failures, design and manufacture for greater safety and economy, and assess operating, maintenance and fitness-for-purpose procedures. (*Failure Analysis Case Studies II* 2001, p.v)

[2] This can be either due to malfunction or failure, since both imply that the artefact stopped performing its function (see 1.5.2).

[3] A spray drier is an artefact often used in mines to dry liquid or slurry fast by means of hot gas.

[4] Engineering Failure Analysis is a journal which "publishes research papers describing the analysis of engineering failures and related studies" (Elsevier 2016).

Although failure analysts start from specific case studies, they do not simply want to explain what happened in this one situation. They also seek knowledge to prevent similar problems in the future. This is expressed in the quote above and also aptly stated by Henry Petroski:

> When failures do occur, engineers necessarily want to learn the causes. Understanding of the reason for repeated failures […] typically leads to a redesigned product. (2001, p.13)

In other words, failure analysts look for ways to use the knowledge about physical causal relations in one specific situation, to draw conclusions regarding causal relations in other situations. These situations range from other instances of the same artefact, over similar artefacts, and even to very different artefacts. One of their goals is furthermore to find ways to alter designs. Their analysis is thus thought to be useful for

(1) understanding (failure of) existing artefacts
(2) altering practices of use of these existing artefacts
(3) designing new artefacts not yet in existence.

Collecting causal knowledge that can be used for these purposes, is not effortless (see also chapter 2). Because their practice deals with malfunctioning artefacts, analysists first need to explain or *diagnose* what happened. Because this often involves new or unseen versions of phenomena, it is often not clear which physical laws or regularities are relevant for this diagnosis. So modelling the failure phenomenon is significantly more difficult than for the cases in the previous chapter. And because the engineers want to use information from one failure to intervene on other artefacts, they need some way of deciding for which artefacts and contexts their knowledge can be useful. This is related to determining the domain of causal knowledge (in contrast to chapter 2, where the domain was stipulated explicitly). In comparison with the previous chapters, failure analysts are confronted with more complex situations of which it is even less clear how they are to be connected to physical theory. I will discuss the relationship between the failure analysis cases and physical theory in chapter 5. Here, I will focus on how failure analysts generalise knowledge about one failure in such a way that it can be used to achieve the three aforementioned goals.

This seems to be an instance of a longstanding philosophical problem regarding generalisation of knowledge from one particular instance or local domain, to other instances or domains. This problem has occurred under many different names and in slightly different forms, including ‘induction’, ‘extrapolation’ and ‘external validity’. Arguably, these problems and debates are similar in the sense that they deal with the question of how to generalise knowledge. I will investigate different types of

generalisation as they occur in failure analysis and what evidence is given/needed for them. Because of the focus on design, the generalisations in failure analysis differ from the more classic examples. What I will call *the design-perspective* sets them apart. Understanding these generalisations can deepen our philosophical understanding of different ways generalisation problems occur and how to solve them. At the same time, it gives us more insight in the complexities of using physical causal knowledge. In section 4.1, I will first discuss induction and extrapolation in more detail. I then pay some attention to the notion of design and clarify what I mean with 'design-perspective'. In section 4.2, I will present three examples from failure analysis practice. They will serve as case studies throughout the chapter.

I will flesh out three distinct types of inference: one that looks like induction, one that looks like extrapolation and one that is still different. I will argue that none of the examples are 'pure' instances of classical generalisations, because they involve artefacts-to-be-designed. They will further specify what is meant with the design-perspective of generalisations. Throughout the chapter I will develop a framework to analyse these inferences. It builds on existing philosophical literature, but I will make suitable adaptations to capture the implications for non-existing artefacts.

In order to analyse the aforementioned inferences, I will first use Cartwright's notion of capacities to present the underlying causal claims and their domains in a standard format. They will allow me to capture the probabilistic nature and locality of the causal claims, while accounting for the stability required for generalisations. Using this standard format, I will clarify the (implicit) inference steps analysts make in their causal generalising reasoning. This will shed light on the evidence required to warrant these steps. Because of the focus on design, we will need a lot of specific information to ensure that recommendations prescribe warranted changes to designs. This is the topic of sections 4.4 and 4.5. There, I will develop my mechanism-based account of the evidence needed to warrant the aforementioned inference steps. It builds on Steel's framework regarding extrapolation in the biomedical sciences. Because the inferences I am concerned with are not strict extrapolations, I will adapt Steel's account, building on sections 4.2 and 4.3. I first determine a mechanism-based criterion of similarity for artefacts. Then, I will define a mechanism-based heuristic to determine when generalisations are warranted in failure analysis. This will also create a clear picture of the required evidence for such generalisations. To finish, I will look back on the tools I used and elaborate why they needed adjusting and developing to guide us in building new artefacts. I will also reflect on the nature of these additions to understand what was missing and how this can be of help in other domains.

Most of the philosophical tools and literature I will use in this chapter originated from philosophical interest in the biomedical and social sciences. In section 1.4, I already

explained why this is not that strange: policy, medicine and technology all direct our focus to use. Technology is furthermore used as a metaphor in many debates in the philosophy of biology and biomedical sciences (e.g. the concept of a function, the idea of a mechanism,…). Clearly, the machine metaphor is illuminating – it has been around for many centuries in some form[5]. But we must be careful not to idealise technology too much. Cartwright and Hardie also write:

> It would be nice if social policy were like a battery. Everything necessary for it to create a current is locked inside the casing; the environment it is to be put to work in is both structured and delimited, like a flashlight or a radio; and there are clear instructions for how it is to be implemented—"Put the end marked *+* here." But for social policies, the requisite scientific and technological knowledge and know – how is often missing. (2012, pp.91-92)

There is a lot of truth to this quote. Yet as will become clear from the examples in section 4.2, making an artefact like a battery work is not as easy as Cartwright and Hardie suggest – let alone adapting it in a successful way.

4.1 Generalisation in failure analysis

4.1.1 Generalisation problems

The previous chapter considered physical laws to be the main source of general knowledge. However, in contexts like failure analysis, engineers are explicitly looking for *causal* knowledge that they can *use* for specific goals. For this reason, they often resort to other sources of collecting general knowledge, viz. knowledge generalisation techniques like induction or extrapolation. In order to frame the current chapter, I will briefly present an overview of several generalisation problems and related philosophical concepts. Taking a closer look at these debates will show that they all share a common concern: the question of knowledge generalisation. More importantly, these debates show that there is no clear consensus about what this question entails or how to solve it.

[5] See part 1 of (*The Routledge Handbook of Mechanisms and Mechanical Philosophy* 2017).

This overview will allow me to develop the discussion of failure analysis more efficiently and show exactly what studying failure cases can teach us.

I will start with induction. The problem of induction has a long-standing history in philosophy. Hume is generally considered the first to draw attention to it:

> As to past *Experience*, it can be allowed to give *direct* and *certain* information of those precise objects only […] which fell under its cognizance: but why this experience should be extended to future times, and to other objects, which for aught we know, may be only in appearance similar; this is the main question. (2007, part 4, §29)

Russell names it as one of the major problems in philosophy:

> It must be known to us that the existence of some one sort of thing, A, is a sign of the existence of some other sort of thing, B, either at the same time as A or at some earlier or later time […]. The question we have now to consider is whether such an extension is possible, and if so, how it is effected. (1912, p.39)

These concerns have not gone unstudied. Philosophers like John Stuart Mill (1843), Charles Peirce (1883) and Rudolf Carnap (1950) paid significant philosophical attention to the problem of induction. Gradually, the problem took on different forms, like the paradox of the ravens (Hempel 1945) and Nelson Goodman's new riddle of induction (1983). The main question, however, still underlies all these on-going debates: the definition of induction is not agreed upon (Vickers 2016), nor has the problem been solved to everyone's satisfaction (Norton 2003). But regardless of the specific definition of (the problem of) induction, these enquiries all engage with the question of how we can justifiably *generalise* knowledge of observed events to unobserved ones.

Another generalisation problem can be found in philosophy of the biomedical and social sciences, specifically in debates regarding *extrapolation*. Steel (2007) introduces the problem of extrapolation as follows:

> Imagine that a chemical […] has been found to be carcinogenic if administered […] in rats, and the question is whether it is also a carcinogen in humans. This is an example of extrapolation: given some knowledge of the causal relationship between X and Y in a base population, we want to infer something about the […] target population. (p.78)

Where the problem of induction posed the question in terms of observed and unobserved events, the extrapolation problem focuses on how to generalise knowledge

between different populations. Illari and Russo (2014, p.48) argue that the widely discussed extrapolation[6] problem is important for both observational and experimental studies, two common methods of gathering data in the biomedical sciences. I introduced these methods in chapter 3 (see section 3.2.3). Recall that observational studies refers to methods where we record observations (Illari and Russo 2014, p.10), while in experimental studies, data are produced "within the experimental setting by manipulating some factors while holding others fixed" (Illari and Russo 2014, p.10).

I do not claim to have presented a complete overview of philosophical literature regarding induction and extrapolation. Yet I hope to have shown that generalisation of knowledge is a significant problem that underlies multiple debates, including debates on induction and extrapolation.

4.1.2 The design-perspective

In this chapter, I will use cases and reasoning from failure analysis to draw attention to another way in which the generalisation problem arises: when we are focusing on creating new things, like artefacts. Both induction and extrapolation deal with events or populations *that already exist*. To be more precise, induction and extrapolation focus on whether our current knowledge generalises to already existing things that we have not studied[7]. These are different questions than whether our current knowledge can help us *to create something new*. As Von Karman famously pointed out, creating new things is central to engineering:

> Scientists discover the world that exists; engineers create the world that never was. (Bucciarelli 2003, p.1)[8]

[6] Illari and Russo also connect it to another topic, namely the problem of external validity (2014, p.18). External validity has been discussed by many philosophers, both formally and informally see e.g. Judea Pearl and Elias Bareinboim (2014), Maria Jimenez-Buedo and Luis Miller (2009), Francesco Guala (2005). It bears striking resemblance to the problem of extrapolation, yet some authors claim it is distinctive (Illari and Russo 2014, p.18). However, this does not matter for the current point.

[7] Naturally, these already existing things can *evolve*, which gives rise to change. Yet this is not the focus of the matter. This evolution is outside our control. I wish to focus on our *creation* of new things.

[8] This quote focuses on the distinction between science and engineering. Recall that Mieke Boon has on several occasions (2011a, 2011b) stressed that we also need to acknowledge engineering science as a scientific practice (see 1.5.8). She contrasts this with engineering practice, which is arguably also the practice Von Karman had in mind.

As I explained in 1.5.3, this is done by designing. It should be clear from that section that there is no agreement about the definition of designing. However, what I want to draw focus to here, is the synthetic nature, which underlies most definitions. As a characterisation, recall the characterisation given by Dorst and Van Overveld:

> Design is a human activity in which we create plans for the creation of artifacts that aim to have value for a prospective user of the artifact, to assist the user in his/her effort to attain certain goals. (2009, p. 456)

In order to create a functioning artefact that performs a certain function, engineers need to combine components in a specific way. As I explained, this is not an easy task and often involves failures and setbacks (see 1.5.3 – 1.5.5). In the context of failure analysis, engineers aim to incorporate the knowledge gained from analysing failed artefacts, into the synthetic activity of designing new artefacts or redesigning old ones. This synthetic activity arguably has "distinctive features of its own" (Kroes 2009, p.406).

The scientific questions that form the core of literature on induction and extrapolation, do not straightforwardly have this distinctive synthetic side. This results in differences: because we synthesise artefacts, we know the designs – we built them – and therefore have more control. Because of their artificial nature, we are faced with less ethical concerns than when we are dealing with organisms. Literature on induction and extrapolation mainly surround scientific questions in fields where we hit more cognitive (we do not have as much knowledge of the organisation of an organism as we have of a human-designed artefact) and ethical limitations (witness the ethical discussions on genetic manipulation or genetic choice). So when focusing on (designing) artefacts, new questions can rise. One of them is how knowledge from known objects guides us in synthesising new ones. Though the traditional problems of induction and extrapolation hold in failure analysis as well, I want to focus on design: on how we create something new using knowledge of current things. This is what I will refer to as the *design-perspective*.

My focus on design is informed by the philosophy of technology, but what I want to achieve with this focus is different. In general, philosophy of technology does not deal with questions regarding knowledge generalisation or what knowledge is needed to make designs possible. As I explained in 1.5.3, philosophers of technology are mainly interested in the practice of designing, whether it is rational, how goals are reconciled etc. Regardless, since I use insights of both disciplines, I hope my analysis can bring the philosophy of technology and the philosophy of science closer together.

4.1.3 Failure analysis as a generalisation problem

Failure analysts proceed from causal claims regarding a specific failed artefact to causal claims regarding other types of artefacts and artefacts-to-be-developed. For their investigation, they benefit from the knowledge provided by designs of the artefact that failed and arrays of lab tests. The resulting causal claims have the form of recommendations and are aimed at altering the processes of use of existing artefacts, or designing new, non-existing artefacts. Unfortunately, contrary to the manuals I described in chapter 2, analysts are often unspecific regarding the domain of their claims – viz. for which artefacts their recommendations hold. Yet, the objects that form the domain of their conclusion, determine what evidence the analysts need to put forward to warrant their conclusion. Whether they know the designs of the intended artefacts, whether they know the context they will be placed in etc. determines what evidence is required. For example, if analysts use one artefact failure to formulate conclusions or recommendations regarding all artefacts of a certain class (e.g. all spray driers), they will need other evidence to warrant their claims than if their conclusion only applies to other artefacts of the same type (e.g. other spray driers constructed according to the same design). Given the differences that can pertain within a certain class of artefacts (they can have different designs, other materials, different functioning, etc.), warranting a claim regarding the entire class is not an easy task. This is not merely a theoretical concern. As will become clear in section 4.2, we can isolate inferences from failure analysts that differ with regard to base and target and therefore require different types of evidence. Yet all the recommendations give us guidance regarding what to do, what to change. Taking the design-perspective, I focus on what recommendations tell us regarding how to combine specific components to create a larger whole with an envisioned function.[9] Given that one of the analyst's aims is to specify design recommendations, failure cases prove significantly insightful to study the way in which current knowledge guides the design of new artefacts.

[9] Recommendations are formulated as normative claims. Nevertheless, like was the case in chapter 2, for these normative claims to be warranted, they need to be based on the right kind of knowledge.

4.2 Three examples of failure analysis as knowledge generalisation

In this section, I will present three examples of failure analysis and flesh out three distinct types of inferences. This will help illustrate what studying these generalisations can teach us regarding the way physical causal knowledge is generalised to help achieve epistemic goals.

4.2.1 The pipe

Talesnick and Baker in "Failure of a flexible pipe with a concrete liner" (2001) present an analysis of a steel sewage pipe with a concrete liner, buried in a clay soil profile, located in Israel. The pipeline never got used because of severe cracking of the inner concrete liner. In their paper the authors want to

> […] determine the cause(s) of damage and the areas responsible. […]. (Talesnick and Baker 2001, p.33)

Talesnick and Baker describe two types of tests: laboratory tests and field tests. In the laboratory test conducted on parts of the pipe, they determined the stiffness, and the vertical deflection or strain, which "induces cracking in the inner pipe line and collapse loads" (Talesnick and Baker 2001, p.34). It was found that

> Severe cracking of the inner liner wall (defined as a crack opening of 0.3 mm […]) occurred at a vertical diametric strain of approximately 1.2%. (Talesnick and Baker 2001, p.34)

This was compared with the measurements made in the field:

> The vast majority of field measured pipe deflections […] exceed the 1.2% limit found to induce severe liner cracking of pipe sections in the laboratory. As a result the extensive damage observed in the internal pipe liner in the field […] is not surprising. (Talesnick and Baker 2001, p.37)

They furthermore argue that

1. Most steel pipes are considered to be flexible and designed accordingly
2. A pragmatic literature based criterion for flexible pipes is a pipe that can withstand a vertical deflection of 2% without damage

3. Though the pipe in question was able to withstand this, the inner liner was not, since it showed cracks at a lower vertical deflection. (Talesnick and Baker 2001, p.37)

One of their conclusions reads:

> ‘‘Flexible” pipes with rigid liners must be designed with care. Flexible pipe design methodologies may be applicable, [provided that] […], the deformation limitations of the liner [are] […] carefully considered. (Talesnick and Baker 2001, pp.42-43)

This is how the reasoning process was laid out in the paper. Let me first attempt to present it in a more logical sequence that draws focus to the different research stages. They say that most steel pipes are considered to be flexible. This entails that they should be able to withstand a vertical strain of 2%[10]. This seems based on engineering knowledge of the authors regarding the properties of steel pipes. They also state that the pipe in question is a borderline case, since it failed under circumstances that would not cause damage to a flexible pipe (2001, p.33). So their reasoning can be presented as follows:

- We assume that flexible pipes have characteristics such that they do not experience damage (viz. retain functional and structural integrity) from strain less than 2%. [assumption]
- Bending tests on pipe segments in the lab show that the inner liner cracked at a vertical strain of 1,2%. Higher deformation can cause cracking. [tests in the lab]
- We measured deformation of more than 1,2% in pipe segments in the field. This deformation is significant when analysing the cracking. [measurement in the field]
- The cracking happened with deformation within the norm for flexible pipes. So even if the pipe itself was correctly designed according to flexible pipe criteria, the inner liner did not perform adequately under the specific circumstances. [inference]
- If we want to use flexible pipe design methodologies in pipes with inner liners, we have to take the strain limitations of the liner into consideration (see quote above). [recommendation/conclusion]

[10] For more information regarding stress and strain, see the appendix.

I represent the base (the artefact which was the subject of the failure analysis) and target (the artefacts they mention in their recommendations) of the inference:

> Base: one pipe
> Target: flexible pipes with rigid liners

4.2.2 The spray drier

In "Creep[11] failure of a spray drier", Carter (2001) presents a failure analysis of a spray drier at the Western Platinum Mine, in Rustenburg, South Africa. A spray drier is an artefact which "dries a finely divided droplet by direct contact with the drying medium (usually air)" in a short retention time (Considine and Kulik 2008, p.5130). The failed spray drier consisted of a cylindrical shell, with an annular gas chamber encircling the base of the shell. Four columns supported the shell. The spray drier suddenly collapsed after 20 years of service while operating normally[12] (Carter 2001, p.73). The aim of Carter's investigation

> […] was to explain the failure and to make recommendations to ensure that it was not repeated on the two remaining driers […]. (Carter 2001, p.73)

The investigation found no significant corrosion, the material was found to be accurately chosen without deterioration (Carter 2001, p.73). Neither were there signs of fatigue, fracture or creep damage. However, there was

> […] clear evidence of a localised buckling deformation in columns and shells in the region of the welded column-shell joint. (Carter 2001, p.74)

Carter's failure analysis methodology consists of comparing "stresses at critical points in the structure with allowable and failure stresses" (Carter 2001, p.75). Carter inferred the allowable and failure stresses from the design code for pressure vessels. He determines the influence of creep conditions on the maximum stress in the structure, both of the column-shell connection and the gas duct. Carter specifies that these calculations are estimates, yet that they "clearly indicate the nature of the failure" (2001, p.77). He calculates that the maximum stress of the structure under creep conditions is

[11] See the appendix for information regarding creep damage.

[12] It is not clear what the author means with "operating normally". Arguably, this is a judgement based on his background engineering knowledge.

significantly above the allowable stress and the failure stress. Based on these findings, he concludes that

> The collapse of the spray drier after 20 years in service is an unusual example of a low stress, high temperature compression creep failure. (Carter 2001, p.77)

Carter measured temperatures in the structure around 300°C, but argues that these are not correct by referring to the design of the spray drier. So Carter argued that the estimated temperature should be considered above 480°C. Even though creep actually redistributes stresses (Carter 2001, p.76) and thus decreases the maximum stress on the structure (compared to the stress values under elastic conditions at 500°C), the resulting stresses were still significantly above the failure stress. According to Carter, this explains the absence of clear creep rupture, and the collapse of the spray drier.
Summarizing the reasoning that led him to this conclusion:

- He assumes that the temperatures inside the shell are higher than measured, based on the working and design of the spray drier. So creep conditions might apply, contrary to what was expected based on the measurements. [assumption]
- Creep distributes stresses, so that the actual stress is less than the calculated elastic stress. For a high creep exponent, the maximum value of stress is about 67% of the maximum elastic stress. [background engineering knowledge]
- In the gas duct, the stress concentration factor under elastic circumstances is 8.9, so this will be redistributed by creep circumstances to 67% of 8.9: 5.9 or, rounding up, 6. [measurement + calculation]
- The stress concentration factor of 6 in creep conditions generates a stress value of 22 MPa. The allowed stress value is 13 MPa, failure arises at 17 MPa. [calculation]
- Since the generated stress value was significantly above the allowed values, this resulted in fractures and collapse of the spray drier. [inference/conclusion]

Furthermore, he makes the recommendation of removing the “lagging and cladding in the region of the annular gas duct and the column-shell joints”, in order to “avoid a similar fate on other more recent (and stronger) spray driers” (Carter 2001, p.77). This refers to the drier, which was “lagged and clad from top to bottom to conserve energy” (Carter 2001, p.73). The established isolation of the drier provides support for his hypothesis regarding the temperature measurements. With this much isolation, he argues that the temperature was probably higher than the value he measured. Similar to

example 1, Carter does not regard the failure of this spray drier as an exception: he makes recommendations regarding other spray driers. He considers his knowledge about what led to the collapse of the analysed spray drier applicable to other spray driers. Yet Carter goes even further in his conclusions. He argues that the purported causal claim also holds for "more recent and stronger" spray driers. The finding that the other spray driers were also lagged and cladded from top to bottom, functions as evidence for this conclusion. The inference I want to draw attention to can be represented in the following way:

Base: 1 spray drier
Target: 2 newer and stronger spray driers with lagging in the same context

4.2.3 The raise boring machine

James describes the failure of a raise boring machine in his article entitled "Catastrophic failure of a raise boring machine during underground reaming operations" (2001). Raise boring is a technique found in underground mining operations, used to produce "interconnecting vertical […] channels (raises) between underground levels in mines" (James 2001, p.159). The process can be characterised by two operations: drilling a pilot hole and back reaming:

> During the pilot hole drilling cycle, drill rods connect the raise boring machine with a bottom-hole assembly consisting of ribbed stabilizers, roller reamer and pilot bit. […] After the pilot hole has been completed, a raise boring head is used to back ream the required raise between the underground levels. (James 2001, p.159)

Back reaming is a technique used to "increase the diameter from that initially drilled" (Slaughter, Cariveau, and Shotton 2006, p.12). So the raise boring machine first drills a smaller channel connecting the two underground levels, after which a reaming head is placed on the machine (on the lower level), which is then pulled back up. The reaming head has a broader diameter than the initial drilled channel and rock cutting abilities, so the pulling up of this head creates a hole of increased diameter compared to the original one.

In his article, James describes the failure of such a machine after 119m of reaming, due to the breaking of all 32 bolts on the raise borer drive (2001, p.160). He describes the site visit, inspection of the fractured bolts and the metallurgic examination consisting of chemical analysis, scanning electron microscopy, optical microscopy and hardness testing. Based on these investigations, he argues that:

(1) The catastrophic failure of the raise boring machine is associated with the fracture of the 32 drive head bolts. Thirty of the bolts have failed as a result of corrosion-induced fatigue.
(2) The bolts have failed due to a combination of high cyclic stressing induced by the operation of the equipment at 13% above maximum thrust and corrosion from the water in the flushing system. (James 2001, p.168)

The two other bolts had failed by 100% tensile overload (James 2001, p.163). He furthermore makes several recommendations:

(1) To prevent corrosion of the bolts the following measures are recommended:
(a) An oil-based red lead primer should be used to create a barrier at the cover-body connection.
(b) Mains water should be used at all times for flushing.
(c) Equipment should not be stored underground for any length of time.
(2) Excessive thrust pressures during operation should be avoided, i.e. the equipment should be used within the limits for which it was designed. (James 2001, p.168)

Contrary to the example above, there is no mention of the specific artefacts these recommendations apply for. His claims appear to include more than a machine nearby. To assess his recommendations further, we need other findings which he mentions throughout the paper:

1. The "centre-bolt" torque was found to be well below the normal figure during dismantling. This could have had the effect of allowing more vertical movement of the drive head cover. (James 2001, p.166)
2. Since the bolts show signs of pitting corrosion, the anti-seize compound with which they are coated "does not afford protection to the surface of the bolts". (James 2001, p.167)

I summarise his reasoning:

- Tests show that the failure was due to fatigue. [tests + background engineering knowledge]
- All fatigue areas of the 30 bolts showed signs of pitting – associated with corrosion. [tests]
- The anti-seize compound with which the bolts were coated thus clearly did not prevent corrosion. [inference]
- The thrust during operating was 13% above maximum, putting greater stress on the weakened bolts. [measurement + inference]

The inference can be represented in this way:

Base: one failure
Target: all (future) bolts in all possible circumstances

Summarizing the three inferences I fleshed out:

(1) this artefact (flexible pipe with rigid liner) to other artefacts of the same type
(2) this artefact (spray drier) to other artefacts with known differences (newer and stronger)
(3) this artefact (raise boring machine) to other artefacts not yet in existence

All of these inferences are related to induction and extrapolation (and external validity), but as I mentioned, they differ in one important aspect: they prescribe, among others, alterations to the design and use of artefacts.

More specifically, the *first* looks like induction specified above, because base and target are of the same type. Yet the result of the inductive step is a design recommendation (how to choose the liner), instead of a general claim about the entire pipe. This is a significant difference with induction as specified above. So this looks like induction, but the focus on redesign differentiates it from the generalisations discussed in section 4.1.1. The result is a prescribed action. The *second* inference is better characterised as a sort of extrapolation, because of the known differences between base and target. Yet the conclusion is again focused on redesign of the target, something that is not represented in the literature on extrapolation. Note that the differences concern properties of the target. They can also take the form of usage or context. This is not surprising given the importance of contexts for artefact behaviour I discussed in 1.5.7. In this case, the context is explicitly stable: the two newer spray driers are located on the same site as the failed one. This is an important part of information contained in the inference and needs to be represented in our analysis. A similar point holds for similar maintenance practices. The *third* inference is similar to the second, only this time, there is no context specified in the recommendations. The inference apparently does not depend on the context the artefacts are placed in. Moreover, it does not merely apply to artefacts that are at the time of the analysis in the vicinity of the analysed artefact, but also to artefacts-to-be-designed. This might seem significantly different from the other examples, but is arguably related.

These inferences are thus arguably slightly different generalisations and all involve design recommendations. In the remainder of this chapter, I will attempt to get a more

profound understanding of whether, why and when they are warranted. These are important questions, since implementing changes in designs is not that straightforward:

> The complexity of many engineered artefacts, together with their interactions with a changing environment, make working out the effects of many design changes either analytically intractable or analytically very difficult [Pavitt, 1984; Nightingale, 2004]. (Nightingale 2009, p.365)[13]

The influence of rather 'small' changes, like whether the centre bolt is torqued to the normal figure (see example 3) can be enormous. Putting together different components in such a way that they combine to realise the envisioned behaviour, is a complex and open-ended process (Dorst and van Overveld 2009, p.456). Making sure that an artefact functions as envisioned therefore often involves "learning, experimentation, testing, and numerous modification and feed-back loops" (Nightingale 2009, p.365). One of the learning occasions is the failure of artefacts. Understanding how knowledge from other, failed artefacts can be implemented in the fickle balance that is created in designing new artefacts, will require different tools than understanding generalisation in scientific contexts that do not focus on this *design-perspective.* Note that the examples also show that the balance is indeed more fickle than suggested by Cartwright and Hardie. Not only can small changes have big consequences, artefacts can also fail after years of functioning normally, like the spray drier in example 2. So it can appear that everything is present for correct and stable functioning, yet after time, in certain contexts, it turns out that something was missed. Knowing how and why artefact failures occurred, when they can occur again and how to repair artefacts and alter designs to prevent failures, is therefore also not a straightforward matter.

[13] In that sense, it is not surprising that the design-perspective has not been excessively studied in the generalisation literature. Many philosophers studying the generalisation problems mentioned in section 4.1.1, focus on the biomedical and social sciences. Because of ethical limitations and cognitive constraints I mentioned already, successfully changing the functioning of an organism is even more difficult than changing the functioning of an artefact.

4.3 Further reflections on the examples

I have showed that these recommendations are not unproblematic. More specifically, they only hold if certain stable causal relations hold. For example, the recommendation

> An oil-based red lead primer should be used to create a barrier at the cover-body connection for all artefacts of type X.

only holds if there is some kind of stability in causal factors across different artefacts and contexts:

> For all artefacts of type X, an oil-based red lead primer is a negative causal factor for bolt breakage.

Intuitively, this comes down to saying that oil-based red lead primers can (and sometimes do) prevent bolt breakage. The term 'causal factor' refers to Giere (see chapter 2). Because of the complexity of the examples in this chapter, and because of my focus on knowledge generalisation, I will need other philosophical tools than the ones from chapter 2 to represent the examples. I will explain the tools that I will use in section 4.4. For now, it is important to see that these causal factors need to be stable in some sense, if they can ever be the base for generalisations. Otherwise, we can never safely assume that they will hold in other situations. This is crucial to understand my case studies, since formulating recommendations from one failure is, as mentioned in the introduction, a type of knowledge generalisation. From the overview in section 4.1.1, it is clear that knowledge generalisation is not without problems. This is also the case in failure analysis. The analysts need to provide reasons *why* their generalisation is warranted. They also need to provide arguments for the generality of their conclusion: the artefacts that they consider part of the domain of the claim determine its validity. To explain this, I adapt example 3. Suppose that James based his analysis on multiple failures of raise boring machines that have the same design, call this type T_1. James then formulates the same recommendation as in the original article, namely

> (a) An oil-based red lead primer should be used to create a barrier at the cover-body connection. (James 2001, p.168)

Suppose that there are raise boring machines with stainless steel bolts (type T_2). These bolts are not susceptible to corrosion in the same contexts as other bolts, so the causal claim does not hold for these machines in similar contexts and correspondingly, the recommendation would be irrelevant. This also shows how we can intuitively understand the stability mentioned above: as related to capacities of (parts of) artefacts in certain

contexts. The stainless steel bolts (in the same context) do not have the increased capacity to corrode. The pipe in example 1 has an increased capacity to bend, while not all other pipes do. I will present elaborate and theoretical underpinnings of capacities in section 4.4. Here, I want to reflect on the need for justification for generalisations, and the required reference to the implied domain.

Steel's discussion (2007) of extrapolation in the biomedical sciences can deepen our understanding of these challenges. According to him, a theory of extrapolation has to solve two basic challenges: the extrapolator's circle and the problem of causally relevant differences between model and target population (2007, p.4). The first

> […] arises from the fact that extrapolation is worthwhile only when there are important limitations on what one can learn about the target by studying it directly. The challenge, then, is to explain how the suitability of the model as a basis for extrapolation can be established given only limited, partial information about the target. (Steel 2007, p.4)

The problem of causally relevant differences, on the other hand,

> […] is a direct consequence of the heterogeneity of populations studied in biology and social science. Because of this heterogeneity, it is inevitable that there will be causally relevant differences between the model and the target population. Thus, an adequate account of extrapolation must explain how it can be possible to extrapolate from model to target even when some causally relevant differences are present. (Steel 2007, p.4)

Based on the discussion of my cases, I argue that both challenges are also present in failure analysis. Analysts need to provide reasons why (1) base and target are similar and (2) account for relevant differences between base and target. Though Steel focuses on heterogeneity in the social and biomedical sciences, causally relevant differences between artefacts are as problematic – witness the spray drier with stainless steel bolts. Recall the previous chapter, where I discussed physical setups as ways to specify the domain of causal claims. So it is important to specify the domain of a causal claim in order to evaluate it. Especially when making design recommendations (see section 4.1.2). To argue that one failure is relevant for other (instances of the same type of) artefacts, analysts thus need to provide evidence. Unfortunately, none of these articles straightforwardly do this. Yet, the recommendations are successful (or so the articles mention). It therefore seems that the inferences are warranted, but the justification is not made explicit or is not reflected upon. Bucciarelli mentions a related observation:

> Epistemological questions about the source and status of engineering knowledge rarely draw [the engineers'] attention. […] If their productions function in accord with their designs, they consider their knowledge justified and true. (2003, p.1)

In the following sections, I will attempt to make the analysts' reasoning and presuppositions explicit. This will allow me to reflect more profoundly on the nature of evidence their generalisations need. For this, I will use Steel's framework (2007) as a starting point. The first step towards this reflection is reframing the recommendations as causal claims.

4.4 Philosophical tools for investigation: making things explicit

4.4.1 Capacities, features and MOD

To reframe the recommendations as causal claims, I first need a more precise definition of the notions 'causal factor' mentioned above. It is inspired by Giere's comparative model of causation (see also chapter 2), but I will connect it to Cartwright's notion of capacities, which is better suited to fit my focus on knowledge generalisation. As noted in section 4.3, causal factors need to be stable in some sense to allow for generalisation. I represent this stability via capacities:

> All bolts have the capacity to break in some contexts.
>
> Corroded bolts have an increased capacity to break.
>
> Therefore corrosion is a causal factor in bolt breaking.

Pointing to causal factors gives engineers a way to indicate capacities, which in turn allow for generalisation. Cartwright illustrates what she means by 'capacity' via the claim that aspirins relieve headaches. According to her, this claim

> […] says that aspirins have the capacity to relieve headaches, a relatively enduring and stable capacity that they carry with them from situation to situation; a capacity which may if circumstances are right reveal itself by producing a regularity, but which is just as surely seen in one good single case. The best sign that aspirins can relieve headaches is that on occasion some of them do. (1994, p.3)

I use Cartwright's notion of capacities for several reasons. One, this notion is inherently connected to probabilistic causal relations (as is causal factor).

> The point is that, for each capacity the cause may have, there is a population in which this capacity will be revealed through the probabilities. (1994, p.121)

Second, capacities are context-dependent; local (Illari and Williamson 2012, p.153). Whether and how a capacity actually manifests, depends on the situation:

> […] capacities […] can be assembled and reassembled in different nomological machines, unending in their variety, to give rise to different laws. (Cartwright 1999, p.52)

Third, the notion of capacity is central to Cartwright's concept of nomological machine:

> […] a fixed (enough) arrangement of components, or factors, with stable (enough) capacities that in the right sort of stable (enough) environment will, with repeated operation, give rise to the kind of regular behaviour that we represent in our scientific laws. (Cartwright 1999, p.50)

On Cartwright's account, nomological machines produce regular behaviour and correspondingly, laws (of nature) only hold "relative to the successful repeated operation of a nomological machine" (1999, p.50). Arguably, all technical artefacts are nomological machines in the sense that they give rise to regular behaviour when functioning properly. Using Cartwright's framework of capacities thus allows for a probabilistic, local notion of causal connection. This is an elegant and useful way to reformulate the recommendations from the failure analysis, though I will make some adaptations.

One adaptation that I need is to reframe them such that the domain of the causal claims is clearly represented (see chapter 2 and the introduction). For the current purposes, we can do this by referring to types of artefacts.[14]

> For all artefacts of type X, c is a positive/negative causal factor for e.

[14] Because I limit my analysis to artefacts, I can refer to types of artefacts to limit the domain. To expand this analysis to natural contexts as well, the notion of physical setup that I defined in chapter 2 can replace artefact types. This does not make a difference for my analysis, but adding it would add to the complexity of this chapter. I therefore chose to leave it to the reader.

Where c and e express a specific value of a relevant variable feature of (part of) the artefact, viz. the value that is thought to be important for the causal claim. I will get back to parts of artefacts later.

Consider example 1. Since Talesnick and Baker analyse only one artefact, their causal claim can be represented as follows:

(E_{F1}) For this pipe, deflection is a positive causal factor for cracking of the liner.

To allow generalisation, (E_{F1}) needs to correspond to:

(E_{F1*}) The deflected pipe has increased capacity for cracking the liner, and it probably manifested for the studied pipe system.

According to Talesnick and Baker, the capacity manifested because the liner cracked and this was at least partly due to the deflection of the pipe. These claims are token level causal claims, referring to one specific artefact. The main causal claims from examples 2 and 3 (and the corresponding capacity claims) can be represented in a similar way.

(E_{F2}) For the Rustenburg spray drier, the lagging and cladding of the annular gas duct is a positive causal factor for collapse of the spray drier.

(E_{F2*}) A lagged and cladded annular gas duct has an increased capacity for collapse of the spray drier and it probably manifested for the Rustenburg spray drier.

(E_{F3}) For this raise boring machine, the corrosiveness of the flushing liquid is a positive causal factor for bolt breaking.

(E_{F3*}) Corrosive flushing liquid has increased capacity for corrosion, which increases the capacity for breaking the bolts and it probably manifested for this raise boring machine.

These are the claims the failure analysts implicitly start from: claims regarding the artefact they investigated. Their recommendations can be represented by referring to types of artefacts:

($E_{F1'}$) Deflection of a pipe is a positive causal factor for cracking of the liner for all artefacts of type X.

($E_{F2'}$) The lagging and cladding of a gas duct is a positive causal factor for collapse of the spray drier for all (?) artefacts of type Y.

($E_{F3'}$) The corrosiveness of flushing liquid is a positive causal factor for breaking of the bolts for all artefacts of type Z.

Analysts go from evidence to causal factor claims. Doing so assumes that by identifying these factors they succeed in identifying some capacity that is stable under certain

circumstances. Clearly, not all pipe liners crack. In order to adequately apply the recommendations (viz. everywhere they might be useful, not where they are thought not to be) we need a clear representation of the circumstances under which the capacity can actually manifest. Cartwright's notion of nomological machine is, as such, not very helpful here. I agree that whether capacities manifest depends on the contexts and the specific machine they are embedded in, which is a big step towards the design-perspective. Cartwright's work has undoubtedly been incredibly important in studying design. But from Cartwright's definition, it is not clear how we can discover which environment and arrangement of components is necessary for specific capacities to manifest in a certain way. Her account does not fully embrace the design-perspective described in section 4.1.2: the actual scientific practice of synthesising components into a functioning whole. So I need to specify how we can discover what the 'right sort of stable environment' and 'arrangement of components' are when designing new artefacts. Based on the examples and the discussion of artefacts I presented in 1.5, I argue that this requires specifying

1. the type of artefact,
2. the relevant causal factor, and
3. the context.

With 'relevant causal factor', I mean the causal factor which corresponds to the increased or decreased capacity. The first two are already present in the current formulation, but the context is not yet represented. Yet, as explained in 1.5, this is an important piece of information. In (re)design literature, this is reflected in the distinction between mode of deployment (MOD) and mechanistic organisation of the artefact (Chandrasekaran and Josephson 2014). The former represents the ways in which the artefact is used, the latter the way the artefact is constructed. For the task at hand, MOD can be understood in a broader sense and also represent certain important aspects of the *environment* of use (e.g. underground, in high humidity) instead of merely *mode* of use (e.g. operating at high thrust):

> For all artefacts of type X and MOD Y, c is a positive/negative causal factor for e.

The way the analysts formulate their recommendations in example 2, seems to imply that they only hold for artefacts in the same context (e.g. a warm climate). In example 3, there is mention of operating *at high thrusts*, but not of requirements for context. MOD can capture both.

4.4.2 Failure mechanisms

So it is clear that failure analysts have certain (un)specified beliefs regarding what factors are relevant to warrant the causal claim. The tools I have presented help to make them explicit. I can now turn to the question of whether and when they are justified. This comes down to determining how to characterise "type X" in the definition above. For this, I will present a two-fold mechanism-based procedure.

Representing failure analysis in terms of failure mechanisms allows me to provide a fruitful answer to the challenges raised in section 4.3: the extrapolator's circle and relevant differences relating to the design-perspective. Moreover, a mechanism-based framework fits well with my characterisation of causal claims in terms of capacities. Cartwright recently connected her notion of nomological machines to the mechanism literature (2009, p.7). According to her, we can understand mechanisms as nomological machines. In this way, her work on nomological machines functions as a connective between capacities (that allow me to express required stability demand) and mechanisms (that can form the basis of the generalisation procedure). Furthermore, the term "mechanism" is often used by failure analysts themselves. For the characterisation, I again use the general definition of a mechanism from Illari and Williamson (2012) that I described in chapter 3 :

> A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon. (p.123)

The examples above all constitute a mechanism in this sense: they refer to entities (e.g. the pipe) and activities (corrode, break), organised in a specific way (the liner covers the inside of the pipe, the bolts hold the driver head in place) such that they are responsible for a specific phenomenon (the failure of the pipe, the collapse of the drier). I find the definition of activities as "producers of change" by Machamer, Darden & Craver (2000, p.3) most appropriate to capture the activities at hand. They are "usually designated by a verb or verb form" and "are constitutive of the transformations that yield new states of affairs or new products" (Machamer, Darden, and Craver 2000, p.4). Clearly, corroding and breaking satisfy this definition. MDC furthermore talk about "bottoming-out":

> Different types of entities and activities are where a given field stops when constructing mechanisms. The explanation comes to an end, and description of lower-level mechanisms would be irrelevant to their interests. (Machamer, Darden, and Craver 2000, p.13)

This is also the case for the verbs I identified as activities in failure analysis: the mechanism of generating stress or corroding is not spelled out in the analyses. These

activities arguably constitute the bottom level in failure analysis. The same holds for the entities involved in their reasoning: analysts do not provide explanations in terms of e.g. atoms or molecules. Finally, because Steel (2007) formulates a mechanism-based strategy to the challenges raised in section 4, I can use his work as a starting point. Characterizing failure analysis in terms of failure mechanisms is thus a promising choice. In the next section, I present Steel's framework before adapting it to fit the failure analysis examples.

4.4.3 Steel on comparative process tracing

The strategy Steel develops is called comparative process tracing (CPT):

> First, learn the mechanism in the model organism, by means of process tracing or other experimental means. Second, compare stages of the mechanism in the model organism with that of the target organism in which the two are most likely to differ significantly. (Steel 2007, p.89)

Steel distinguishes two steps to CPT. The first (process tracing or other experimental means) deals with mechanism discovery.[15] In failure analysis this often has to do with background engineering knowledge, such as the properties of flexible pipes. I will not discuss this further.

The second step is relevant for the current purposes: look for significant differences between model and target. If significant differences pertain, we cannot be sure that the behaviour of the model will be replicated in the behaviour of the target:

> Significant differences are those that would make a difference to whether the causal generalization to be extrapolated is true in the target. (Steel 2007, p.89)

To check for such differences, we need "generalisations asserting that objects of specified types resemble one another in certain ways though not necessarily in others" (ibid.). Knowledge of these generalisations allows us to (1) check in a directed way whether the specific differences occur and correspondingly (2) judge whether the extrapolation is warranted or not. Steel developed his framework for the biomedical and social sciences, so he is looking for generalisations like

[15] Process tracing refers to two strategies for discovering mechanisms: schema instantiation and forward chaining/backtracking. For more strategies of mechanism discovery, see (Darden 2017).

> Features A, B, and C of carcinogenic mechanisms in rodents usually resemble those in humans, while features X, Y, and Z often differ significantly. (Steel 2007, p.89)

Because of my focus on to-be-designed artefacts and design recommendations, Steel's generalisations will not do the job. The comparisons Steel suggests are apt to capture biomedical examples, but are too unspecific to reflect the amount of control we have over, and knowledge we possess of, artefacts. Since we have more knowledge of and control over many artefacts (specifically the ones I am talking about), we *can* compare existing to non-existing artefacts in a more specified and detailed way than Steel allows in his framework: we can compare designs on specific points for example. But more importantly, because of the difficulties of adapting designs (see section 4.2), we *need* to be specific and warrant the applicability of recommendations thoroughly. Uninformed or unspecific implementation of recommendations can have unforeseen consequences; small changes can result in grave problems. Moreover, though organisms evolve too, Steel's account is not focused on the required knowledge to *actively change* organisms[16] – he does not take the design-perspective. Yet this is exactly what I am interested in regarding artefacts. So like with Cartwright, Steel's CPT provides a good basis to model generalisations to non-existing artefacts, but needs some changes and additions.

In the following sections I will show how Steel's framework can be adapted to fit failure analysis. I will first elaborate on what it means for artefacts to be similar, and present a mechanism-based account for that. I then proceed to adapt and apply Steel's CPT to fit the failure analysis examples.

4.5 A mechanism-based generalisation framework

4.5.1 Similarity

CPT depends on knowledge of likely similarities and dissimilarities between base (e.g. mice) and target (e.g. humans). But before the actual CPT can begin, we need to ensure that there is enough similarity between base and target to allow the base to function as

[16] There is the question of genetic modification which I mentioned in section 4.1.2. I am not focused on this scientific practice, but I believe my account (and the adapted version of CPT that I present) has the potential to be useful in this context as well. I get back to this in the conclusion.

a suitable model for the target. In biomedical sciences, this comes down to knowledge of some mechanism e.g. the metabolism in mice and humans. I argue that, in failure analysis, this comes down to knowledge of whether the failed submechanism of the artefact is present. It refers to a submechanism[17] which helps sustain the artefact and its functioning. In the example of the raise boring machine, the submechanism (M) that failed is the mechanism attaching the cover of the boring head to the body. It can be characterised as follows:

> Entities: connecting parts, cover, body
> Activities: connecting parts immobilise cover, cover is immobilised
> Organisation: cover is fastened onto body via connecting parts

Every artefact containing a submechanism of this type is a candidate for the domain of the recommendation. Note that this implies that different mechanisms containing submechanisms of the same type can be part of the domain. Remembering the characterisation of causal claims underlying recommendations I discussed in the previous section, the relevant question we need to ask is:

> For all artefacts containing a submechanism of type M and operating in MOD Y, is c a positive/ negative causal factor for e?

In the following section, I answer this question by adapting CPT.

4.5.2 CFPT – Comparative failure process tracing

I already discussed mechanistic evidence in the previous chapter. As I signalled there, mechanistic evidence can also be invoked to extrapolate causal knowledge. By adapting Steel's CPT to fit the generalisations from my examples, I will shed light on how information about the existence of a mechanism provides evidence for generalising physical causal knowledge. Recall that Steel (2007) developed his framework mainly to deal with organisms. Compared to organisms, there is a crucial difference to analysis of artefacts: we know the designs – we built them. This allows for greater manipulability. Where Steel, for the biomedical examples, has to refer to "knowledge of likely dissimilarities", we can be more specific with regard to the nature of these dissimilarities

[17] Lindley Darden describes a similar process in relation to mechanism discovery, viz. modular subassembly (2017, p.261).

in failure analysis. I argue that for the failure mechanism[18], we need 3 comparison points and a check for counteracting mechanisms. I will first discuss the 3 comparison points. They are

(1) the types of parts,
(2) the organisation, and
(3) the activities and corresponding properties.

The connection between activities and properties fits my definition of activities in the MDC sense:

> […] activities determine what types of entities (and what properties of those entities) are capable of being the basis for […] acts. Put another way, entities having certain kinds of properties are necessary for the possibility of acting in certain specific ways, and certain kinds of activities are only possible when there are entities having certain kinds of properties... (Machamer, Darden, and Craver 2000, p.6)

It furthermore ties in with the capacities that ensure the stability of the causal factors. So this is a fruitful connection. Call the combination of our 3 comparison points and counteractive mechanism checking 'Comparative Failure Process Tracing' (CFPT). I will now illustrate these comparison points via the examples in order to facilitate arguing for each of them.

Revisiting the pipe

This is an interaction between instances of two types of parts: a pipe and a liner. They are organised in a specific way: the liner covers the inside of the pipe. They furthermore interact such that the pipe deflects and the liner responds by cracking. This interaction is connected to specific properties of the pipe and the liner: the pipe is flexible, the liner is rigid.[19] There is no MOD specified – Talesnick and Baker argue that the pipe-liner interaction is due to a design problem. Representing these aspects in a standard format:

[18] The similarity criterion referred to a submechanism partly responsible for artefact functioning. The failure mechanism is responsible for the failure phenomenon. The functioning and the failure of the artefact are two different phenomena, and thus call for different mechanisms. So the (sub)mechanisms that was (partly) responsible for the functioning of the artefact, is not the same mechanism that is responsible for the specific artefact failure. However, the mechanisms can overlap.

[19] There are threshold values, but this does not matter here. There is no reference to threshold values in the recommendations of the failure analysts.

Types of parts involved: pipe, liner
Properties: pipe is flexible, liner is rigid
Organisation: liner covers inside of pipe
MOD: no usage or context specified

In a similar way, consider the spray drier.

Revisiting the spray drier

Types of parts: cladded and lagged shell of a spray drier, gas duct
Properties: the shell has a mass, the gas duct can break
Organisation: the shell leans on the gas duct
MOD: no specified usage, creep temperatures

And finally, I revisit the raise boring machine.

Revisiting the raise boring machine

Types of parts involved: parts that immobilise cover of the drive head, flushing liquid
Properties: flushing liquid is corrosive, immobilisation parts are susceptible to corrosion
Organisation: flushing liquid engulfs immobilisation parts
MOD: high thrust, no specified context

With this in mind, I can argue that these points can house significant differences.

The aspects of CFPT

Types of parts: Suppose we know that a specific raise boring machine does not use cooling liquid (T3). The type of part (cooling liquid) is not represented. Therefore we cannot say that the failure mechanism will also take place in T3 raise boring machines. There is no entity to partake in the activity that is crucial to the failure mechanism (viz. corroding the bolts). In this case, there are even reasons to believe the mechanism will not be active.

Properties: As MDC stated, entities partake in activities because of some properties - they need to have the capacity associated with the causal factor specified in the causal claim underlying recommendation. Even if all entities are present in the target artefact, they need to have the required properties to take part in the relevant activities. I already mentioned one example in section 4.3, when talking about a raise boring machine with

stainless steel bolts rather than bolts that can corrode. Another example would be a spray drier with an unbendable and unbreakable gas duct. It will not be susceptible to the same failure mechanism as is present in example 2, since the gas duct cannot break. The shell of the spray drier still has the capacity to break gas ducts, but it will not manifest because the gas duct does not have the capacity to break. If the liner of the pipe is not made of concrete, but instead of some other flexible material, it will not crack.[20]

Organisation: This is fairly straightforward. If the organisation of the entities differs significantly, the failure mechanism will not be active. If the shell does not rest on the gas duct, it will not generate stress on the duct and the same mechanism can therefore not be said to hold. Other mechanisms can of course be present that generate the same effect, but I would argue that they need other recommendations.

Summing up: So these three points of comparison describe features where significant differences can pertain. Note that MOD also remains an important point of comparison, but is arguably distinct from the other points, since they deal with aspects of the failure mechanism. If no dissimilarities are found, the failure mechanism can be active. Combining CFPT with the information above, we arrive at the following characterisation of the recommendations' domain:

> For all artefacts that (1) contain a submechanism of type M, (2) are used in MOD Y and (3) pass the CFPT, c is a positive/negative causal factor for e.

Let me briefly discuss the required check for counteracting mechanisms. If, for example, the shell of the spray drier had ventilation holes while being lagged and cladded, the failure mechanism might also not be active. Determining what counts as a counteracting mechanism again depends on a lot of background knowledge and applications of multiple scientific regularities. To illustrate, consider a submechanism that is placed in a new artefact, but behaves completely different there; in an unforeseen way. Based on the discussion in section 4.2, this is a real possibility. This means that there are causal relations that we have not taken into account. Specific parts, properties, features of the organisation or specific counteracting mechanisms have not been considered in designing the new artefact. This 'failure' of the submechanism teaches us about new causal relations, about mechanisms that we did not consider to be counteracting, about

[20] The phrasing of example 1 confirms this point: they make explicit reference to flexible pipes with rigid liners, implying correctly that their claim does not necessarily hold for non-flexible pipes and/or non-rigid liners.

connections that we considered negligible but weren't. Designing an artefact is attempting to make a nomological machine, a complex system which behaves as we want it to and not differently (most of the time). If something does not behave the way we envisioned, we missed something in the description or shielding. My framework allows the engineers to specify what happened and why, instead of just acknowledging something somewhere went wrong – which is arguably important to make warranted decisions and act in warranted ways. Yet this is no plea for infallible designs. I agree with Bucciarelli that

> […]there will always be a potentially problematic state of affairs not considered, overlooked, unimagined, unconstructed, no matter how many safety procedures one invokes or how imaginative and free wheeling your brainstorming session about possible contexts of use may be. (2003, p.30)

There can always be aspects that haven't been taken into account. So looking for a procedure to provide a definitive answer regarding whether a specific failure will occur, is a futile undertaking. I have therefore not described an *algorithm* for determining the domain of failure recommendations, but rather presented a *heuristic* to determine whether recommendations are relevant. As I mentioned already, we often have specific, reliable and direct ways to check the features mentioned in the heuristic for specific artefacts: designs. Information regarding types of parts, properties of these parts, organisation and possible counteracting mechanisms are often mentioned there. So by representing the mechanisms that failure analysts identify as responsible for the failure phenomenon in a way that highlights these comparison points, engineers can actually learn from past failures in an easy way and thoroughly check whether the recommendations are relevant for their specific situations. In biomedical sciences, CPT informs us which model population will succeed most in capturing the mechanism in the target:

> Thus, comparative process tracing yielded the conclusion that the rat was a better model than the mouse. (Steel 2007, p.91)

The target population in biomedicine is often humans. In failure analysis, on the contrary, what the model teaches us determines the target; the domain of the analyst's recommendations.

4.5.3 Relation to Cartwright and Steel

Now that my framework is completely spelled out, I can further specify why Steel's and Cartwright's notions did not suffice to capture how we generalise to non-existing

artefacts, and correspondingly, how we can use failure of existing artefacts to create new things. As I mentioned in section 4.4.1, Cartwright's capacities allow for a local, probabilistic notion of causality. This was very useful to characterise the causal claims from the failure analysis examples. Yet Cartwright's notion of capacities (and the related notion of nomological machines) as such cannot characterise when inferences to non-existing artefacts are warranted. Capacities and nomological machines are not specific enough for this goal. Consider the example of the aspirins:

> The best sign that aspirins can relieve headaches is that on occasion some of them do. (Cartwright 1994, p.3)

Yet when we want to design a new type of aspirin, we need to know the specific circumstances under which they relieve headaches, so that we can ensure that the newly designed aspirin will also manifest this capacity. Clearly, we need knowledge of capacities for this, but that is not all, we need more. Besides the capacities, we also need specific information of the environment and arrangement of components needed to make the capacities manifest. Only then can engineers (or chemists) attempt to successfully synthesise components into a larger system with a specific function – which is what it means to design an artefact (recall sections 1.5 and 4.1.2). Cartwright's discussion of nomological machines touches on this (e.g. (Cartwright 1999, p.64)), yet does not give us specific guidance as to how we should collect or present this information. My framework, on the other hand, gives insight into the nature of the knowledge required by the design-perspective and facilitates its presentation via a mechanism-based procedure. It is not surprising that Cartwright's account cannot answer the questions I am concerned with. Her main goal is to refocus the debate on laws to capacities by arguing for "a patchwork" of laws, instead of a pyramid. She developed 'nomological machines' for this goal. As such, it is not sufficiently specific to guide the specific question of how we can develop new artefacts from failed ones.

A similar point holds for Steel. When focusing on how we succeed in designing new things, we need significantly more information than merely reference to likely similarities and differences in the operating mechanisms. As I argued in section 4.5.2, we need certain specific things to stay the same: the parts, their relevant properties, the organisation, the mode of operation. It is of utmost important to specify these comparison points if we want to understand why failure analysts can make recommendations regarding objects-to-be-designed. Above that, we need a way to specify the role background engineering knowledge plays. So all of this 'messiness' cannot be captured by referring to similar mechanisms and likely differences. In creating new things, our control is greater, but the amount of required evidence to warrant generalisations, is as well.

Conclusion

In this chapter I focused on evidence for generalising physical causal knowledge regarding artefacts and on the difficulties determining the domain of such more generalised knowledge. I argued for my third specific claim (III): contrary to what was the case in the previous chapters (and is often assumed throughout philosophy of physics, see 1.1.4 and 1.2.1), universal or fundamental laws are not the only source of general knowledge. We do not just attempt to fit phenomena under laws. In many domains, like failure analysis, local knowledge is generalised to fit other contexts and help achieve other goals. Focusing on contexts where this happens brings my analysis another step closer to scientific practice and, unsurprisingly, adds more complexity as well. Let me recapitulate what I showed.

I started with an overview of several problems and related on-going debates regarding knowledge generalisation. Reflecting on engineering practice and specifically failure analysis, I have argued that philosophical discussions of such problems need to be expanded to cope with creation of new artefacts. In general, discourse on knowledge generalisation focuses on targets already in existence. I argued that certain reasoning (specifically relating to the designing of new artefacts) in scientific practices, including failure analysis, is not adequately characterised in this way. Yet there are several 'benefits' to studying artefacts: we often have greater knowledge of artefacts, since we designed them. This is especially the case for the artefacts I focus on. This greater knowledge, combined with less ethical restraints due to their artificial nature, results in greater control over them. Because of that, new questions arise.

One of them, the question I focused on, asks how we can use knowledge from existing artefacts to design new ones. In other words, how can knowledge of (failed) artefacts guide us in combining functional components into a lager system with an envisioned overall function? I called this the design-perspective on generalisation. Such generalisations are present in, among others, failure analysis. I have provided a first attempt to characterise these inferences and reflect on when they are justified. I illustrated this with case studies from failure analysis. I fleshed out three different types of inferences to new artefacts: one that looks like induction, one that looks like extrapolation and one that is neither. I proceeded to analyse these inferences by representing them in a standard format based on Cartwright's notion of capacities. This allowed for probabilistic, local causal claims, while accounting for the stability required for generalisations.

Because of my focus on design, I adapted Cartwright's discussions on capacities and nomological machines. In order to successfully build nomological machines (what

artefacts are), we need more information than a general reference to components with stable capacities and a right sort of stable environment; we need to know what the components are and what the 'right sort' of environment is. We need qualitative information and a way to represent it. Only then can we create an artefact with the envisioned functional behaviour. I also argued that we needed to specify the mode of operation, to account for different types of use and contexts the artefact can be placed in. Combining these insights, I argued that engineers implicitly look for claims of the following format:

> For all artefacts of type X and MOD Y, c is a positive/negative causal factor for e.

Recall that the mode of operation (MOD) also included the context in which the artefact was placed (see also 1.5.7). I then presented a heuristic to determine which artefacts belong to 'type X' – the domain for which the inference is valid and what evidence we need to determine this. For this, I used and adapted Steel's mechanistic framework of warranted extrapolation. It hooked nicely onto the mechanistic representation of artefacts I presented. It depends on "likely similarities and dissimilarities of base and target". Like with Cartwright, my focus in artefacts and design demanded adapting Steel's framework. Because of the specific synthetic nature of designing and the complexity of changing designs, I argued that we need more specific information to determine when recommendations are warranted for artefacts-to-de-designed. Fortunately, we also have more knowledge of artefacts, so we can provide this information. Starting from these insights and the examples from failure analysis, I argued that we can develop a more specific description of what it means for artefacts to be similar or different in ways relevant to the inference. Regarding similarity, I argued that (new) artefacts are candidates for the domain of the inference if they contain the submechanism which failed in the original artefact. I represented this in the following way:

> For all artefacts containing a submechanism of type M and operating in MOD Y, is c a positive/ negative causal factor for e?

I then developed a mechanistic (heuristic) procedure to check for relevant differences and thus determine (non-deterministically) a justified answer to this question. As mentioned, the artificial nature of 'artefacts' allowed for greater specificity than the cases Steel deals with. I argued that to determine whether the study's failure can also manifest for a certain artefact, we need to check three points of comparison, viz. whether relevant *parts* are present, whether these parts have the appropriate *properties* and whether they are *organised* in a way that is similar. Finally, we need to check for

counteracting mechanisms. I called comparing the three aforementioned points and checking for counteracting mechanisms "Comparative Failure Process Tracing":

> For all artefacts that (1) contain a submechanism of type M, (2) are used in MOD Y and (3) pass the CFPT, c is a positive/negative causal factor for e.

I stressed that all of these steps require a great deal of background engineering knowledge and that this procedure should therefore not be seen as an algorithm, but merely as a tool for making the inferences explicit. In this way, I hope to have provided a first attempt to reflect on generalisations that deal with artefacts not yet into existence.

The account presented in this chapter is not only relevant for philosophical reflections on physical causal knowledge. Because it draws attention to an underinvestigated aspect of knowledge generalisation (viz. when and how can we generalise in order to design new objects), my analysis can possibly provide inspiration for similar inferences in other innovation contexts – such as genetic manipulation and pharmacology. If medical practitioners want to engineer new drugs or chemical compounds based on knowledge we possess today e.g., they also need strategies to determine when and whether our current knowledge provides a base to warrant new designs. Moreover, by understanding the differences between technical scientific practices and social and biomedical ones, we can gain a more profound understanding of these sciences and their relations. My analysis is also relevant for engineers. For one, it allows failure analysts to present stronger arguments for their recommendations by making the required evidence explicit. My framework can even provide ways to make the analyst's recommendations more precise. By using my framework analysts can tie their formulations more clearly to the evidence that other engineers can use to evaluate the whether the recommendations are relevant for the machine and context the engineers are interested in. This is related to the importance of packaging knowledge in a way that allows travel.[21] The same can be said about representing information from failure analysis cases in such a way that engineers can reuse them in other contexts, to avoid failure or make design adjustments. Moreover, other engineers that wish to use the knowledge from failure case studies need lots of background knowledge of a specific artefact or failure sub-discipline to evaluate whether a case study is relevant for their situation. Above that, many sub-disciplines in (design) engineering have their own methodologies and criteria (Dorst and van Overveld 2009, p.456). The procedure specified in this chapter, combined with the

[21] See Sabina Leonelli's work (for instance (2010)) for detailed discussions on the importance of packaging in the biomedical sciences.

tools to make the level of generality explicit, can aid analysts and designers in determining whether failure scenarios are applicable to their specific cases.

However, the main merit of this chapter is the role it plays in this dissertation. The cases I discussed in this chapter come from the actual practice of failure analysis, and the difference in complexity with the cases from chapter 2 and 3 is noticeable. For one, the failure analysis cases involve causal knowledge regarding both explanation (the diagnosing of what happened in the analysed artefact) and intervention (formulating recommendations to prevent similar failures). This is due to their focus on applying causal knowledge to benefit technological development. So while I conceptually separated these two contexts of use in chapter 2 and 3, in reality explanation and intervention are often intertwined. A second important point is the added complexity that these cases show regarding generalising knowledge and determining the domain of this knowledge. Failure analysts actively look for causal knowledge that can be generalised to a certain broader group of artefacts. This shows that universal laws are not the only source of general knowledge, far from it. Correspondingly, this more general causal knowledge that analysts produce is often still rather local. And gathering this knowledge is all but easy, let alone determining the domain of this knowledge. While in the previous chapters, I took these aspects as rather stable and unproblematic, looking at scientific practice that collects and uses physical causal knowledge shows that they are very wonky and precarious. As such, the analysis in this chapter reinforces my main point that using physical causal knowledge is a topic worthy of philosophical investigation. And the more we study actual scientific practice, the more complicating factors arise and the more important it becomes to reflect on these practices. In this way, this chapter contributes to realising my two generic aims (A) and (B), by identifying yet another interesting philosophical topic related to using physical causal knowledge.

Because of their precarious, the generalisations of failure analysis would be hard to capture by a framework in terms of laws. My framework in terms of capacities works better for this purpose, since the causal knowledge remains local and we can reflect on the appropriate circumstances of the capacity to manifest. In the special sciences, similar problems lead to the rising prominence of reasoning in terms of mechanisms.

Something I mentioned in this chapter (see 4.1.1) but did not really pay attention to, is the fact that these cases from failure analysis do not line up with physical theory that easy. As you may have noticed, the analysts did not use fundamental laws of physics to model the failure phenomenon. In the next chapter, I will show that this can be related to the enormous amount of more local physical regularities that are being used and developed in the engineering sciences. However, this invites the question of how they relate to the laws of physics, and how to choose between all these regularities. In the next chapter, I will present a pragmatic answer, focusing on use.

References

Boon, Mieke. 2011a. In Defense of Engineering Sciences: On the Epistemological Relations Between Science and Technology. *Techne* 15 (1):49-71.

———. 2011b. Two Styles of Reasoning in Scientific Practices: Experimental and Mathematical Traditions. *International Studies in the Philosophy of Science* 25 (3):255-278.

Bucciarelli, Louis L. 2003. *Engineering philosophy*: Delft University Press.

Carnap, Rudolf. 1950. *Logical foundations of probability*. Chicago, IL, US: University of Chicago Press.

Carter, Paul. 2001. Creep Failure of a Spray Drier. In *Failure Analysis Case Studies II*, edited by D. R. H. Jones. Amsterdam: Pergamon.

Cartwright, Nancy. 1994. *Nature's Capacities and Their Measurement*: Oxford University Press.

———. 1999. *The dappled world: essays on the perimeter of science*. Cambridge: Cambridge University Press.

———. 2009. How To Do Things With Causes. *Proceedings and Addresses of the American Philosophical Association* 83 (2):5-22.

Cartwright, Nancy, and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford, New York: Oxford University Press.

Chandrasekaran, Balakrishnan, and John R. Josephson. 2014. Function in Device Representation. *Engineering with Computers* 16 (3-4):162-177.

Considine, Glenn D., and Peter H. Kulik. 2008. *Van Nostrand's scientific encyclopedia*. Hoboken, N.J.: Wiley.

Darden, Lindley. 2017. Strategies for discovering mechanisms. In *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, edited by S. Glennan and P. Illari. London: Taylor & Francis.

Dorst, Kees, and Kees van Overveld. 2009. Typologies of design practice. In *Philosophy of Technology and Engineering Sciences*, edited by A. W. M. Meijers, D. M. Gabbay, P. Thagard and J. Woods. Amsterdam: North Holland.

Failure Analysis Case Studies II. 2001. Edited by D. R. H. Jones. 1 edition ed. Amsterdam: Pergamon.

Goodman, Nelson. 1983. *Fact, Fiction, and Forecast, Fourth Edition*. 4th Revised ed. edition ed. Cambridge, Mass: Harvard University Press.

Guala, Francesco. 2005. *The methodology of experimental economics*: Cambridge University Press.

Hempel, Carl G. 1945. Studies in the Logic of Confirmation (I.). *Mind* 54 (213):1-26.

Hume, David. 2007. *An enquiry concerning human understanding*: Oxford University Press. Original edition, 1748.

Illari, Phyllis McKay, and Jon Williamson. 2012. What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science* 2 (1):119-135.

Illari, Phyllis, and Federica Russo. 2014. *Causality: Philosophical Theory meets Scientific Practice*. 1 edition ed: Oxford University Press.

James, Alan. 2001. Catastrophic failure of a raise boring machine during underground reaming operations. In *Failure Analysis Case Studies II*, edited by D. R. H. Jones. Amsterdam: Pergamon.

Jimenez-Buedo, Maria, and Luis M. Miller. 2009. Experiments in the social sciences: The relationship between external and internal validity.

Kroes, Peter. 2009. Introduction to part III. In *Philosophy of Technology and Engineering Sciences*, edited by A. W. M. Meijers, D. M. Gabbay, P. Thagard and J. Woods. Amsterdam: North Holland.

Leonelli, Sabina. 2010. Packaging Small Facts for Re-Use: Databases in Model Organism Biology. In *How Well Do 'Facts' travel*, edited by P. Howlett and M. Morgan: Cambridge University Press.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. Thinking about Mechanisms. *Philosophy of Science* 67 (1):1-25.

Mill, John Stuart. 1843. *A System of Logic, Ratiocinative and Inductive, being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation*: John W. Parker.

Nightingale, Paul. 2009. Tacit Knowledge and Engineering Design. In *Philosophy of Technology and Engineering Sciences*, edited by A. W. M. Meijers, D. M. Gabbay, P. Thagard and J. Woods. Amsterdam: North Holland.

Norton, John D. 2003. A Material Theory of Induction. *Philosophy of Science* 70 (4):647-670.

Pearl, Judea, Elias Bareinboim, and others. 2014. External validity: From do-calculus to transportability across populations. *Statistical Science* 29 (4):579-595.

Peirce, Charles Sanders. 1883. A theory of probable inference. In *Studies in logic by members of the Johns Hopkins University*. New York, NY, US: Little, Brown and Co.

Petroski, Henry. 2001. Success and failure in engineering. *Practical Failure Analysis* 1 (5):8-15.

The Routledge Handbook of Mechanisms and Mechanical Philosophy. 2017. Edited by S. Glennan and P. Illari. London: Taylor & Francis.

Russell, Bertrand. 1912. On the notion of cause. *Proceedings of the Aristotelian society* 13:1-26.

Slaughter, Robert H., Jr., Peter T. Cariveau, and Vincent W. Shotton. 2006. Back reaming tool. Smith International, Inc.

Steel, Daniel. 2007. *Across the Boundaries: Extrapolation in Biology and Social Science*: Oxford University Press.

Talesnick, Mark, and Rafael Baker. 2001. Failure of a flexible pipe with a concrete liner. In *Failure Analysis Case Studies II*, edited by D. R. H. Jones. Amsterdam: Pergamon.

Vickers, John. 2016. The Problem of Induction. In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta.

Appendix

Here I present extra information regarding certain engineering terms relevant for my cases in chapter 4. The information will also be useful for chapter 5. I focus on stress, strain and creep.

Stress is applied force per unit area:

> The force per unit area is called the stress. (Feynman, Leighton, and Sands 2011, II-38-2)

Amount of stress is often expressed in terms of the Stress Concentration Factor:

> Any abrupt change in the cross section of a loaded component causes the local stress to increase above that of the background stress. The ratio of the maximum local stress to the background stress is called the stress concentration factor, or SCF for short. (Ashby and Jones 2012, p.266)

Strain is deformation that happens as a result of stress (Ashby and Jones 2012, p.34). More specifically, it is "stretch per unit of length" (Feynman, Leighton, and Sands 2011, II-38-2). It represents change in dimensions (Mitchell 2004, p.380).

Two types of strain are distinguished, depending on whether deformation is reversible: elastic (temporary) strain and plastic (permanent) strain (Ashby and Jones 2012, p.117):

> almost all materials, when strained by more than about 0.001 (0.1%), do something irreversible: and most engineering materials deform plastically to change their shape permanently. (ibid)

The region in which deformation is reversible, depends on the material:

> To this point, we have limited the discussion to small strains—that is, small deviations from the equilibrium bond distance, such that all imposed deformations are completely recoverable. This is the elastic response region, one that virtually all materials possess […]. What happens at larger deformations, however, is dependent to some extent on the type of material under consideration. Beyond the elastic region, we enter a realm of nonrecoverable mechanical response termed permanent deformation. There are two primary forms of permanent deformation: viscous flow and plastic flow. […] Other types of nonelastic

> responses to applied forces, [include] fracture and creep, […]. (Mitchell 2004, pp.380-390)

To understand what creep entails, we need to know that the nature of deformation changes with temperature:

> At room temperature, most metals and ceramics deform in a way that depends on stress but which, for practical purposes, is independent of time:
>
> $$\varepsilon = f(\sigma)$$ elastic/plastic solid
>
> (Ashby and Jones 2012, p.311)

When temperatures rise, deformation becomes dependent of time. This process is referred to as "creep" or *viscoplasticity*. It is defined as time-dependent deformation under constant stress, usually at elevated temperatures.

> As the temperature is raised, loads that give no permanent deformation at room temperature cause materials to creep. Creep is slow, continuous deformation with time: the strain, instead of depending only on the stress, now depends on temperature and time as well:
>
> $$\varepsilon = f(\sigma, t, T)$$ creeping solid
>
> (Ashby and Jones 2012, p.311)

Creep gives rise to creep strain:

> A typical creep experiment involves measuring the extent of deformation, called the creep strain, ε, over extended periods of time, on the order of thousands of hours, under constant tensile loads and temperature. (Mitchell 2004, p. 432)

Finally, the creep exponent:

> By plotting the log of the steady creep rate, ε_{ss} against log s at constant T, […] we can establish that
>
> $$\varepsilon_{ss} = B\sigma^n$$
>
> where n, the creep exponent, usually lies between 3 and 8. This sort of creep is called "power-law" creep. (Ashby and Jones 2012, p.315)

References

Ashby, Michael F., and D. R. H. Jones. 2012. *Engineering Materials 1: An Introduction to Properties, Applications and Design*. 4 edition ed. Amsterdam: Butterworth-Heinemann.

Feynman, Richard P., Robert B. Leighton, and Matthew Sands. 2011. *The Feynman Lectures on Physics, boxed set: The New Millennium Edition*. Slp edition ed. New York: Basic Books.

Mitchell, Brian S. 2004. *An Introduction to Materials Engineering and Science for Chemical and Materials Engineers*: John Wiley & Sons.

Chapter 5 Epistemic authority: a pragmatic approach

In this fifth chapter, I shift my attention to my final topic, viz. the relation between laws and useful physical causal claims. I do this by examining why a certain physical causal claim or a piece of physical causal information is given what I call *epistemic authority*. With epistemic authority, I mean that the piece of information is trusted to base interventions or explanations on; that it is trusted to reach epistemic goals. Traditionally, this authority has been linked to lawhood: fundamental laws deserve this authority. In this chapter, I will show that when we focus on using physical causal knowledge, this view is untenable. In this way, this chapter constitutes an argument for my final specific claim, viz. (IV). I will show that from a use-perspective, the focus on fundamental laws does not aid our understanding of scientific practice. Correspondingly, this chapter identifies a final interesting philosophical issue related to useful physical causal knowledge and concludes my exposition in favour of my two generic aims (A) and (B).

I will first discuss two traditional ways in which philosophers have attempted to define laws: one in terms of necessity and one in terms of an epistemic mark. I will then reflect on a case from the previous chapter (viz. the collapsed spray drier) with regards to the regularities that the engineer used to diagnose the failure. I will specifically focus on one regularity, called "the Neuber rule" and attempt to explain why it has what I will call *epistemic authority*. I will argue that a necessity approach to laws cannot explain why the Neuber rule has epistemic authority. For this part, I will build on arguments by Mathias Frisch. I will then show that by expanding the idea of epistemic mark as criterion for laws with insight from Sandra Mitchell, I can explain why the Neuber rule has epistemic authority. This will result in an approach to laws that is more in tune with scientific practice. At the same time, it explains how engineers work successfully with so many different regularities with different degrees of generality and necessity.

This chapter will also scrutinize the central position of laws in the philosophy of physics. I will show that not only are physical regularities not enough to reach epistemic

goals, the physical regularities we need are often not the laws of physics. This is related to the idea that the (fundamental) laws exhaust the content of physics. These ideas were not actively present in the cases I discussed (except in chapter 3), but they are present in philosophy of science and philosophy of physics in particular (see chapter 1). My cases so far have been compatible with the assumptions: we might have other ways of getting general knowledge, like the generalisation in chapter 4, but one can still consider the (fundamental) laws of physics as main sources. Because of the central position that laws have taken in philosophy of physics and in philosophy of science (see 1.1.3), these assumptions are quite tough and deserve separate attention. In this chapter, I scrutinise and argue against them. As such, I will show that the assumption that studying the laws of physics suffices to provide philosophical insight in all domains of physical scientific practice, including how we use physical causal knowledge, is untenable. I will reflect on my main argument in the conclusion of this dissertation.

Introduction

In chapter 1, I mentioned the focus on laws in the philosophy of science (see 1.1.3). I connected this to the central position that laws take in philosophy of physics, and to the view that physics is an exemplary science. Yet the focus on laws is undermined in the special sciences, since the generalisations of the special sciences are thought not to be lawlike. Here, I will reflect in more detail on the philosophical discussion on lawhood in general and I will connect this to the debates on laws in the specific sciences.

Consider the following question:

> Why does this metal rod lengthen when heated?

According to physics, this is due to the laws of thermal expansion. When a question can be answered by invoking "a law of nature", in many contexts, there can be no discussion. Laws of nature are seen as irrefutable, as the way things are. Laws are also considered special: not all regularities are laws. Bas Van Fraassen presented a much-cited example referring to the size of spheres:

> All gold spheres are less than a mile in diameter.
> All uranium spheres are less than a mile in diameter. (1989, p.27)

The first universal sentence is not considered to be a law, the latter one is. This distinction is thought to be important, because laws are seen as capable of performing

functions that accidental generalisations are not. This is motivated by the observation that, throughout the sciences, laws are given what I will call *epistemic authority*: they are trusted as guides for epistemic activities such as prediction, explanation and manipulation. The latter two are the ones I have focused on throughout this dissertation. While I have attempted to show that using physical knowledge to achieve epistemic goals is not that easy, philosophers have generally agreed that laws allow us to achieve them. An example is given by Van Fraassen:

> The moon orbits the earth. Why does it do so? What reason can there be for expecting it to continue to do so? (1989, p.183)

The reason is the law of gravitation (Van Fraassen 1989, p.32). Many philosophers (though not Van Fraassen himself) think that it is precisely *laws* that allow us to explain e.g. why the moon orbits the earth and make us successful in our prediction that it will do the same tomorrow.

Especially explanation has often been connected to laws. Cartwright discusses this via the distinction between phenomenological and theoretical laws (1983, p.1). Phenomenological laws describe what happens. But theoretical laws explain: they are not merely about appearances, but about the reality behind them (Cartwright 1983, p.1).[1] The idea that there is a special connection between laws and explanation has also influenced the philosophical analysis of scientific explanation. Think for instance of Hempel's covering-law approaches, such as the deductive-nomological model (or DN-model). In this model, a potential explanans[2] is characterised as follows:

> (DN) The ordered couple (L, C) constitutes a potential explanans for the singular sentence E if and only if
> (1) L is a purely universal sentence and C is a singular sentence,
> (2) E is deductively derivable from the conjunction L&C, and
> (3) E is not deductively derivable from C alone.
>
> (Weber, Van Bouwel, and De Vreese 2013, p.2)

In the formulation above, 'L' refers to a law[3]. By deductively subsuming the phenomenon to be explained under a law, an explanation is given. As Cartwright argues, this is not

[1] Cartwright herself does not agree and argues that it is the causes that explain, not the laws.

[2] . In general, an explanation is considered to consist "of an explanandum E (a description of the phenomenon to be explained) and an explanans (the statements that do the explaining)" (Weber, Van Bouwel, and De Vreese 2013, p.2).

[3] In some explanations, more than one law may play a role.

limited to Hempel's account. Patrick Suppes's probabilistic model of causation, Wesley Salmon's statistical relevance model, and even Bengt Hanson's contextualistic model, all rely on the laws of nature (Cartwright 1983, pp.44-45). The pragmatic account of explanation that I used in chapter 3 also relied on laws – but recall that laws were everything but sufficient. However, in the literature, the emphasis has been put on laws. So for many philosophers, if we can show that a phenomenon can be derived from a law (together with other information), it is explained. Laws and explanation are intimately connected.[4]

A similar story can be told for other epistemic goals, such as manipulation and prediction. In the quote above for example, Van Fraassen asked what reason we have to expect that the moon will orbit the earth again tomorrow. The answer was "the law of gravitation". This reason behind the uniformity in nature ensures that we can make predictions regarding the behaviour of the moon:

> That there is a law of gravity is the reason why the moon continues to circle the earth. […] and if we deny there is such a reason, then we can also have no reason for making that prediction. We shall have no reason to expect the phenomenon to continue, and so be in no position to predict. (Van Fraassen 1989, p.32)

So according to many philosophical views, we can a.o. explain and predict phenomena in a successful and warranted way by referring to laws. Laws ensure that our epistemic undertakings are trustworthy. This is what I mean when I say that laws receive epistemic authority in scientific practice and philosophical analysis. This is not a novel point. Nelson Goodman already said it wasn't a novel point back in the 80's (1983, pp.20-21). So laws are used predictively. Note that it should be clear from the previous chapters that the story is not that simple. However, what I want to focus on here, is why exactly laws are granted (and should be granted) this epistemic authority, unlike *accidental generalisations*. Philosophers have attempted to legitimate this in several ways, often by defining what makes a law a law. I will discuss two strategies: necessity and epistemic mark. I start with necessity.

[4] As Holly Andersen notes, many philosophers have argued that laws fail in fulfilling these roles and started focusing on mechanisms (2017, p. 158). I already discussed this in chapter 3. I agree with her point, but as I argued in 1.1.4, laws still take up a central position in philosophy of physics. See also Meinard Kuhlmann's contribution to the same handbook (2017).

5.1 Legitimating epistemic authority

5.1.1 Necessity

A first traditional way of distinguishing between laws and accidental generalisations has to do with modal power: something in nature is thought to necessitate the truth of laws, while this is not the case with other regularities (Carroll 2016, §3). This is often expressed in terms of necessity (Van Fraassen 1989, p.28) and connected to the ability of laws (in contrast to accidental generalisations) to support counterfactuals (Psillos 2002, pp.145-146). The main idea behind this way of legitimating the epistemic authority of laws, is that what follows from the laws is, in some sense, necessary.

There has been a lot of debate regarding how to understand this necessity[5], yet the specific strength does not really matter here[6]. So I will adopt the weakest form of necessity as described by Van Fraassen: the intensionality of the laws. Intensionality of laws expresses that if "it is a law that A" is true, then "A" is necessarily also true[7] (Van Fraassen 1989, p.29). So even philosophers who admit that laws can be contingent generalisations, often still see them as necessitating the truth of their consequences. Note that this is not an epistemological criterion, but a metaphysical one: it is not what we know of the laws, their functioning etc. that warrants a belief in A, but the fact that the law is *there.* This switch from epistemological activities to legitimation in terms of metaphysical modality is subtle, but it is there and it seems to simply happen.

There is of course a rationale behind this. If the laws express what is "precluded or allowed by nature" (Beatty 1995, p.239), they are sensible guides in the sense that they discourage us from trying something that is precluded by nature. The laws tell us the rules of nature's game. If we know these rules (or 'reasons', as Van Fraassen says), we can exploit them to arrive at specific goals. In this way, there seems to be a strong connection between the regularities expressing necessities and their being trustworthy

[5] David Armstrong, Frank Dretske and Michael Tooley (hereafter ADT) are notable defenders of laws as necessity relations between universals (Psillos 2002, p.163). See also Armstrong (1985), Dretske (1977), Tooley (1977).

[6] One topic of discussion is for example whether the laws themselves are necessary. In the weakest interpretation I am taking here, the laws themselves need not be necessary. They could have been different, but given that they are what they are, they necessitate their consequences.

[7] Many philosophers argue that this notion is far too weak to capture our intuitions regarding what it is to be a law. I am interested in scientific practice and not in intuitions, therefore I will not discuss this.

guides in epistemic activities. Yet, like I described in chapter 1, this idea has been undermined in some of the special sciences, for instance in biology (see 1.1.3). Thinkers like John Beatty (1995),Stephen Gould (1990) and Martin Carrier (1995) have argued that evolution could have given rise to other regularities in biology than the current ones. Because of this, they claim, many regularities in biology "do not express any natural necessity" (Beatty 1995, p.239) and should not be called laws. Not everyone agrees with a definition of lawhood in terms of necessity. For instance, philosophers who are interested in epistemic questions regarding laws, like myself, are not keen to accept a very metaphysical characterisation of laws. In the next section, I will discuss an alternative, which focuses on epistemic authority.

5.1.2 Epistemic mark

Another attitude towards this epistemic authority of laws has been to *equate* lawhood with epistemic authority. This strategy, which Psillos called "the epistemic mark" (2002, p.141), can be captured as follows:

> It is a law that all Fs are Gs if and only if (i) all Fs are Gs, and (ii) that all Fs are Gs has a privileged epistemic status in our cognitive inquiry. (Psillos 2002, p.142)

This view was defended by among others, Nelson Goodman (1983), Richard Braithwaite (1953) and Alfred Ayer (1963). Their accounts do not make reference to metaphysical properties of laws, but they also do not give us any understanding of why specific regularities are used in cognitive inquiry. Moreover, this strategy has been criticised for being too subjective and anthropomorphic (Psillos 2002, p.142). Mill (1911), Ramsey (1928) and Lewis (1973) (hereafter MRL) are often put in the same boat. They defend an approach dubbed "web-of-laws approach" (Psillos 2002, p.148), where laws are the axioms in the systematic organisation of our knowledge. According to Lewis, a regularity is a law

> if and only if it appears as a theorem (or axiom) in each of the deductive systems that achieves a best combination of simplicity and strength" (73, 1973 in Psillos 2002, p.149)

Though this approach made lawhood more objective than other epistemic mark accounts, many philosophers still hold that it is too subjective or anthropomorphic (Psillos 2002, pp.155-157).

This is not meant as an in-depth overview of the debates surrounding laws, and many excellent works on this topic are readily available[8]. For my current point, this brief summary is enough to frame the problem of epistemic authority and the ways it has been handled in the past. The debate regarding laws has still not reached a consensus. In the philosophy of biology, the topic is even mostly abandoned. Craver and Kaiser even see it as outlived and suggest we should instead focus on how generalisations in biology can play the role they do:

> Nobody anymore denies that there are stable regularities that afford prediction, explanation, and control of biological phenomena. Whether such stable regularities count as laws depends on what one requires of laws, but it is undeniable that generalizations of this sort do many kinds of work in biology. What remains is the admittedly difficult work of showing how this is possible. (Craver and Kaiser 2013, p.127)

I agree with the suggested shift of focus, but there is more to be said. For one, which specific regularities provide a secure basis for explanation, prediction,… is not agreed upon by philosophers of science. And second, one of the reasons philosophers have sought a definition of laws, is to explain *why* certain regularities (viz. laws) can be used for prediction and explanation, and certain other regularities (viz. accidental regularities) cannot. So shifting the focus of the debate is not as easily done as Craver and Kaiser suggest. Yet in this chapter, I will make an attempt.

I will investigate *how* regularities can play the roles they do in scientific practice, without posing the question in terms of what defines laws. This involves arguing that neither of the approaches mentioned (viz. necessity and epistemic mark) can successfully account for why certain regularities get epistemic authority in science. For one, focusing on metaphysical criteria for lawhood does not increase our understanding of scientific practice. Consequently, if scientific practice is what we are interested in, as I am, focusing on metaphysical markers for lawhood will not be useful. Especially because the regularities that are the best candidates for having this mark, are not often used in scientific practice. I will get back to this in 5.2. Second, merely referring to our epistemic attitude towards certain regularities is not very insightful. I will get back to this in 5.3. I instead propose a shift in focus away from attempting to provide a definition of laws. I will focus on what we successfully do with regularities and what this can tell us about those regularities. So contrary to many philosophers, I start from the observation that

[8] See for instance Van Fraassen (1989), John Carroll (1994), Marc Lange (2000), Carroll (*Readings on laws of nature* 2004), Psillos (2002).

certain regularities have epistemic authority and attempt to understand what this tells us about why regularities are used in science. In that sense, I am following the approach of Goodman, Ayer and Braithwaite, but I am developing it. To show what this other perspective can teach us, I will take a closer look at the engineering sciences. I will first present a case study from this domain. This will help to get a better understanding of what the engineering sciences are and how they successfully use regularities. These cases will be the guiding sources of information for understanding the epistemic authority that regularities get in scientific practice.

5.2 Epistemic activities in the engineering sciences

5.2.1 What are the engineering sciences

In 1.5.8 I already introduced Boon and Knuuttila's definition of the engineering sciences as striving "through modelling to explain, predict or optimise the behaviour of devices, processes, or the properties of diverse materials, whether actual or possible." Recall that Boon also argued that the engineering sciences have a very distinctive modelling practice, which is not reducible to physics (2011, p.64), nor to technology or design (Boon and Knuuttila 2009, p.1).

Technology and engineering are related however, since "much research in the engineering sciences is aimed at creating and understanding physical phenomena that may be put to technological use" (Boon 2011, p.66). Yet as I already explained, the relationship between these domains (fundamental physics, the engineering sciences and technology) is underinvestigated and does not benefit from the traditional debate regarding laws. If laws are considered to receive their epistemic status from the metaphysical necessity they express, only research regarding regularities that express necessity seems legitimate. So I believe the alternative, practice-engaged understanding of why certain regularities are trustworthy guides for predictions, explanations, etc., will benefit philosophical reflections on the engineering sciences.

To get a better grip on what it is that engineering scientists do, I will revisit the case of the failed spray drier from the previous chapter (see 4.2.2). Contrary to what I did in that chapter, I will now focus on one of the regularities that Carter used to explain the failure of the spray drier, and on not the recommendations that he made. I will argue that a specific rule Carter uses for his analysis, namely the Neuber rule, gets epistemic

authority, though it is not straightforwardly a law according to the two defining approaches I described in 5.1.

5.2.2 A creepy case

Recall that Carter diagnosed that the failure of the spray drier was a result of creep, which referred to non-elastic (viz. irreversible) deformation due to stress at high temperatures. Based on his investigation, he recommended removing the "lagging and cladding in the region of the annular gas duct and the column-shell joints", in order to "avoid a similar fate on other more recent (and stronger) spray driers" (Carter 2001, p.77). To understand Carter's analysis, I repeat some information about creep that I also mentioned in the appendix of the previous chapter. Under normal temperatures

> most metals and ceramics deform in a way that depends on stress but which, for practical purposes, is independent of time:
>
> $$\varepsilon = f(\sigma)$$ elastic/plastic solid9
>
> (Ashby and Jones 2012, p.311)

When temperatures rise, deformation becomes dependent of time. This process is referred to as "creep" or *viscoplasticity*

> As the temperature is raised, loads that give no permanent deformation at room temperature cause materials to creep. Creep is slow, continuous deformation with time: the strain, instead of depending only on the stress, now depends on temperature and time as well:
>
> $$\varepsilon = f(\sigma, t, T)$$ creeping solid
>
> (Ashby and Jones 2012, p.311)

Creep gives rise to creep strain; deformation of the material:

> A typical creep experiment involves measuring the extent of deformation, called the *creep strain, ε*, over extended periods of time, on the order of thousands of hours, under constant tensile loads and temperature. (Mitchell 2004, p.432)[10]

[9] *Stress* (σ) is applied force per unit area (Feynman et al. 2011, II 38-2). *Strain* (ε) is deformation that happens as a result of stress (Ashby & Jones 2012, p.34). More specifically, it is "stretch per unit of length" (Feynman et al. 2011, II 38-2). Two types of strain are distinguished, depending on whether deformation is reversible: elastic (temporary) strain and plastic (permanent) strain (Ashby & Jones 2012, p.117).

A final important concept is the creep exponent.

> By plotting the log of the steady creep rate, ε_{ss} against log s at constant T, […] we can establish that
>
> $$\varepsilon_{ss} = B\sigma^n$$
>
> where n, the creep exponent, usually lies between 3 and 8. This sort of creep is called "power-law" creep. (Ashby and Jones 2012, p.311)

Carter specifically makes use of a Neuber calculation to determine the creep stress concentration factor in notches, which states that

> the product of shear stress and shear strain concentration factors of a notched body of nonlinear material was independent of the external load level. (Härkegård and Sørbø 1998, p.224)

The Neuber calculation is often mentioned in this form:

$$K_t^2 = K_\sigma K_\varepsilon$$

with

$K_t =$ theoretical stress concentration factor

$K_\varepsilon = {}^{\varepsilon_e}/_{\varepsilon_{ne}}$ actual strain concentration factor

$K_\sigma = {}^{\sigma_e}/_{\sigma_{ne}}$ actual stress concentration factor

Another way of understanding the Neuber calculation is the following:

> In connection with low-cycle fatigue analysis, where inelastic strains are generally confined to the notch root area, Neuber's (generalized) rule implies that the product of equivalent stress and strain, $\sigma_e\varepsilon_e$ is equal to the same product under linear elastic conditions. (Härkegård and Sørbø 1998, p.224)

Carter uses the Neuber rule (together with design information of the spray drier) to determine the actual stress in the column-shell from the value of stress and strain under elastic (viz. reversible) conditions.

Let's focus on this Neuber calculation. Since Neuber presented his rule in 1961, it received a lot of attention. For one, people found that in certain conditions, the rule overstates the stress. It has also been used in different ways, attempting to model different circumstances and materials. This practice, as well as the article by Carter, shows that the Neuber rule has epistemic authority: it is used to explain phenomena

[10] Note that this is not a quote by the philosopher Sandra Mitchell, but Brian S. Mitchell, professor in the department of chemical and biomedical engineering at Tulane.

(e.g. the failure of the investigated spray drier) and predict phenomena (e.g. the behaviour of some more recent and stronger ones). Recall that Carter's article was published in a specialised failure analysis journal, and reprinted as part of a reference set of real failure investigations" (*Failure Analysis Case Studies II* 2001, p.v). The goal is to communicate findings to other engineers and engineering students (*Failure Analysis Case Studies II* 2001, p.v). So the engineering sciences community accepts this kind of explanation and prediction, based on regularities like the Neuber rule. Moreover, the articles (like Carter's) often mention that their recommendations (based on predictions) are successful.

How can we legitimate the fact that engineers trust the Neuber rule to make explanations and predictions? Coming back to the debate I sketched in the introduction, one option is to show that it is a law – since laws are thought to rightly receive epistemic authority. In section 5.1, I discussed two criteria that have been proposed for lawhood: expressing necessity and having an epistemic mark. Let's first consider necessity: does the Neuber rule express any? One straightforward difficulty is that the Neuber rule is not without exceptions, since it overstates stress in some situations. This is not compatible with the intensional view on laws sketched in the introduction. If we can find an exception (say an observation of $\neg A$), then "it is a law that A" cannot be true, since we can derive "A" from this, which would yield a contradiction with our observed$\neg A$. This suggests that the Neuber rule is not a law in this sense and correspondingly, that we have no reason to believe that we can warrantedly extrapolate it to new contexts. Like Van Fraassen said, if there is no reason for the regularity, we cannot trust it to make predictions. Yet it *is* being trusted and used to make predictions. What does this entail for the validity and authority of the rule? It does not look good regarding necessity. Yet there are laws that are not universally valid, such as Ohm's law, which is not valid for e.g. diodes. Maybe one could still save the Neuber rule by showing that it does express some necessity, yet has a limited domain. One might want to show that it is derivable from laws that we think certainly express necessity, specifically the laws of physics. This is the topic of section 5.3.

After discussing necessity, I will get back to the epistemic mark in section 5.4. As I have shown, engineers have the epistemic attitude associated with laws towards the Neuber rule. So according to e.g. Goodman's account, the Neuber rule counts as a law[11], and laws can be trusted to make predictions and explanations. This can be seen as a step in the right direction, but what does this actually tell us? It tells us very little. We trust

[11] Note that MRL might disagree.

the regularity, so we can trust it. Though this is a proper beginning for an epistemic account of why we trust regularities, it is not informative enough. I will expand this argument in section 5.4.

5.3 The laws of physics: the real deal

I will now turn to the option to ground the Neuber rule by showing that it is deducible from more fundamental regularities that are thought to be necessary. Someone advocating this strategy hopes that, since the Neuber rule is a deductive consequence of laws that do necessitate their consequences, the necessity will carry over to the Neuber rule. The most obvious candidates for such fundamental laws are the laws of physics. This might be seen as a solution for all the local generalisations that I described in the previous chapter as well. In this section, I will investigate the assumptions on which this line of reasoning rests and raise problems for two of them. In section 5.3.2, I will use arguments from Mathias Frisch to scrutinise the idea of fundamental laws. I will then expand his arguments regarding modelling to argue that, even if there are fundamental laws, there is no guarantee that their necessity will make any difference for the modal power of the Neuber rule. This is the topic of 5.3.3.

5.3.1 Grounding the Neuber rule

An important observation regarding the possibility of reducing the Neuber rule to more fundamental (and necessary) laws of physics, is that up until now, no attempt has been successful. From the engineering literature on creep, we can conclude that there is no model of creep in terms of fundamental laws – let alone of creep in notches which is crucial to this case. This is not for a lack of trying. Frank Nabarro (2004), for instance, attempted to build such a fundamental model . Nabarro was one of the pioneers of dislocations in solids (Brown 2010, p.275) and spent a significant part of his life studying various modes of creep and having one specific type of creep in crystals named after him (viz. Nabarro-Herring creep), only two years before he died. In his article from 2004, Nabarro discusses an overview of the different models of power law creep. Power law creep, or steady-state creep, refers to the third stage of creep (after initial rapid extension and primary creep) which occurs before rupture (Nabarro 2004, p.659). This is a particularly interesting stage for engineers (Nabarro 2004, p.659), and is also the type

of creep present in the spray drier case study. Throughout the decades, many models have been presented and three are often cited: two models by Weertman and one by Spingarn and Nix. These models attempt to explain why, for a wide range of stresses, "the creep rate is […] closely proportional to a power of the stress, with an exponent of about 4.5-5.0" (Nabarro 2004, p.661). Each model starts from a different idea regarding the mechanism of creep (viz. by glide of dislocations of a certain density until they are blocked by other dislocations, by glide of two dislocation configurations, or by dislocation glide on a single slip system in each grain (Nabarro 2004, p.660)), resulting in different predictions of the creep rate. Nabarro argues that every model

> either lacks physical probability or fails to predict a rate of creep of the order which is observed experimentally". (2004, p.660)

So a definitive or accepted theoretical model of creep has not been developed yet. The Neuber rule is not an exception in this matter. Many other regularities that are constantly used in the engineering sciences have not been theoretically grounded yet.

But science is a gradual endeavour and we might at some point in the future succeed in deriving the Neuber rule from thermodynamical laws. Should we keep on using the Neuber rule counting on the fact that it will at some point be grounded? To answer this question, let me look more closely at what is happening here. The idea is that the laws of physics apply to all possible phenomena, and we just need to figure out how. This is related to the prestige physics, especially fundamental physics, has: it is seen as the most mature science, a role model for other sciences (see for instance (Norton 2003)). Correspondingly, the laws of physics are often considered as exemplars of laws of nature: universally valid, necessary, fundamental[12]. According to this view, the Neuber rule is, like all other regularities about the physical world, simply shorthand for a model in terms of more fundamental laws. As a result, on this view, all the rule's properties (including any necessity it expresses) are due to its relation to the fundamental laws.

This way of reasoning actually rests on many assumptions regarding the laws of physics and modelling practices. Specifically, it rests on the ideas that (1) the laws of physics express necessity and (2) are fundamental, which in some sense implies that they can capture all phenomena. And regarding modelling practices, there is the assumption that (3) necessity carries over through modelling practices. Though this final assumption

[12] This is obvious from the debates surrounding whether biology has laws, mentioned in the introduction. In analysing the nature of regularities in biology and debating their law-likeness, laws of physics were often used as a contrast class or point of reference (see (Beatty 1995), (Brandon 1997), (Carrier 1995)).

is less easy to recognise, I will argue in section 5.3.3 that it is in fact there. I will not deal with the first assumption, but will show that (2) and (3) are not as unproblematic as often assumed. In the following section I will first scrutinise the purported fundamental nature of the laws of physics. For this part, I will build on Mathias Frisch's arguments against foundationalism developed in the context of causal reasoning in physics.

5.3.2 Frisch on the laws of physics

I already explained Frisch's plea for a broader approach to philosophy of physics in chapter 1. For clarity, I will briefly repeat the relevant aspects here. Recall that Frisch wants to broaden the debate on causation in physics with insights from scientific practice. He argues that, contrary to what many other philosophers of physics assume, the fundamental equations in isolation do not cover the entire representational content of a theory. Instead, we need to take into account the user and the context to fully understand the representational content of a theory, even for physical theories. His alternative account of scientific practice in physics starts from a "pragmatic and structural account of representation" (Frisch 2014, p.37). In chapter 1, I explained that this means that we need to take the user and the context into account to understand how physical representations of phenomena work. This is because we cannot reach all the epistemic goals we want to via laws or equations only. We need to build *models*. The

> […] physical processes that interact with the production and annihilation of elementary particles are not, and in fact cannot be, modeled quantum field theoretically. Instead physicists use resources from theories such as classical electrodynamics, fluid dynamics, and solid-state physics to model the causal structure within which the quantum-field theoretic interaction is embedded. (Frisch 2014, pp.80-81)

It is the modelling practices that really matter. This lines up well with my current purpose: understanding how regularities get epistemic authority. Part of the epistemic authority of the regularities is that they are being used to model phenomena (see also 3.1.2).

Frisch uses this pragmatic account of representation to formulate a convincing argument against what he calls scientific foundationalism. This is meant to capture the view that

> physics aims to discover fundamental micro theories that have a universal domain of application and in principle possess models of all phenomena. (Frisch 2014, p.37)

Frisch shows that scientific foundationalism is inherently incompatible with a pragmatic account of representation, like the one he defends. His argument draws from physical modelling practice and is pretty straightforward:

> [C]ontrary to what the foundationalists assumes [sic], we do not have fundamental models representing macroscopic phenomena. To actually construct a quantum-mechanical model of a macroscopic body of water, we would have to solve the Schrödinger equation for on the order of 1025 variables – something that is simply impossible to do in practice. (Frisch 2014, p.38)

Because Frisch defends a pragmatic account of representation, a hypothetical solution to the Schrödinger equation for a body of water, does not qualify as a model of the body of water. In order for something to count as a representation of a system, it needs to be used to represent that system or some other system sufficiently close to the one we want to represent. So if we successfully model a glass of lemonade with the Schrödinger equation, this is a pretty good indicator that the equation can also be used to represent a glass of water. Yet a hypothetical solution, or a model for a very different system cannot provide this indication. His point holds for all the so-called fundamental laws. Note that it is not clear which laws should be seen as fundamental. Sandra Mitchell gives us a pragmatic characterisation:

> It is not clear that anything that has been discovered in science meets the strictest requirements for being a law. However, if true, presumably Newton's Laws of Motion, or The Laws of Thermodynamics, or the Law of the Conservation of Mass/Energy, would count. (2002a, p.330)

Frisch's argument goes through regardless of any specific set of 'fundamental' laws: looking at current physical modelling practices, no laws are effectively used to represent all phenomena, so no set of laws is really fundamental. Instead, macro phenomena need to be represented by macro theories. Our

> […] putatively fundamental micro theories do not represent higher-level macro phenomena […] (Frisch 2014, pp.24-25)

I have shown that Frisch lays out what is, in my opinion, a strong argument against the view that the fundamental equations of (theoretical) physics can (1) represent all the phenomena we are interested in and (2) are all we need to understand scientific practice, and I have supplemented his arguments with regard to epistemic authority. This strengthens the points I made in the previous chapter. In scientific practice, modelling is what matters, and not only are the laws not enough to build these models and use them, not all phenomena (especially not macro phenomena) can be modelled with the

fundamental equations. Simply because a steam engine is thought to behave according to the laws of thermodynamics, does not mean it can be represented by referring to the ideal gas law and the conservation of energy principle.

Frisch's conclusions do hinge on the acceptance of a pragmatic account of representation, and people may disagree with this – and they have. But Frisch does not merely posit this view, he defends it by showing how important the role of modelling is in physical practice. Moreover, since I am explicitly interested in scientific practice, I agree that focusing on modelling and actual representations is a legitimate choice. Only representations and models that are actually being used (or were used at some point) in science to represent certain phenomena, can be part of scientific *practice*. And only by studying these can we understand how we successfully use physical knowledge for reaching epistemic goals. So a pragmatic view of representation is actually quite suited for understanding scientific practice. And if we take scientific practice seriously and accept such a pragmatic account of representation, there can never be a set of laws that 'in principle' represents all possible phenomena, not even laws that express necessity. Laws only capture phenomena via models, and only those phenomena for which models have actually been built and are used in epistemic practices (or phenomena close enough to those). Frisch specifically wants to draw attention to this neglected part of physics (viz. the modelling practices) and argues that they are an important part of the scientific practice. Because of this, a pragmatic account of representation fits best.

So if Frisch's arguments go through, attempting to show that the Neuber rule can be deduced from fundamental laws will not help us in warranting its epistemic authority, since there are no real fundamental laws that can actually be used to model everything.

Because I am specifically focused on the engineering sciences, it is noteworthy that Frisch's findings go against a tendency in philosophy of physics – a tendency that has negative influences on the understanding of the engineering sciences:

> [T]he picture of science that arises is that, in the end, a complete knowledge of the fundamental laws and/or building-blocks presents us with knowledge from which everything else can be deduced, and therefore makes any other epistemic practice intellectually empty. (Boon 2011, p.64)

Boon has criticised this tendency in her defence of the engineering sciences. Though it might be useful in some contexts to try to reduce models of complex phenomena (like artefact behaviour) to more fundamental laws, Boon (2006) shows that this view ignores a big part of actual modelling. Her arguments cohere with those made by other philosophers in different contexts, such as Cartwright (1983).

5.3.3 The problem of modelling

As I said above, Frisch's arguments are convincing, but controversial. Yet even if his arguments do not go through and there are indeed some fundamental laws, there is another complication for trying to warrant the Neuber rule's epistemic authority in the candidate fundamental laws. Here I will show that even if we hold that a set of purported fundamental laws is able to cover all phenomena, there still is no guarantee that the models that we build with them express the same necessity.

The candidate fundamental laws, regardless of what they are, will be used to build models, and those models will guide us in epistemic activities like manipulation and prediction. This at least complicates the connection between necessity and epistemic authority of laws, since there is a 'layer' of modelling between the phenomenon and the law. If it is to be necessity which grants regularities their epistemic authority, the necessity needs to be something that is not damaged by the modelling practices. Yet there is nothing necessary about the way physicists model phenomena. I already introduced the philosophical debate on modelling in chapter 3. I emphasised that the way we construct a model for a phenomenon depends on what we believe is relevant for that phenomenon. Frisch's arguments can be seen as expanding this view: depending on the interests of the user and depending on the context, different choices will be made, resulting in different models. The models do not follow from the laws in any necessary way.

> To the extent that resemblance plays a role in representation, it does so as a function of the representation's use. For example, in certain contexts we identify a representation's target with the help of selective resemblances between representation and target. Yet which aspects are important in assessing the likeness between representation and target is given by the context in which the representation is used. (Frisch 2014, p.28)

Frisch makes this point as part of his pragmatic framework, but does not focus on it since he is mainly interested in causation.

I use this point to develop another difficulty for grounding the epistemic authority of the Neuber rule in candidate fundamental laws. Once we acknowledge that we only use fundamental laws via models in our manipulations and predictions, and that those models do not follow necessarily from the laws, warranting regularities like the Neuber rule is not as unproblematic as it seems. Because in order to warrant the rule, scientists need to model the phenomenon described by the rule (viz. creep in notches) via more fundamental laws (a.o. laws of thermodynamics) and derive the Neuber rule from this model. Yet modelling a phenomenon is dependent on the user and the circumstances. In

order for necessity to warrant the epistemic authority of a regularity, the necessity has to be something that is undamaged by the modelling practices. But there is no necessity to the practices, so it is hard to see how this would work or why this would be the case. So even if there are fundamental laws, the chance that necessity will actually carry over through the modelling is really slim. Simply assuming that it will carry over is actually not taking modelling practices seriously and again would not be in spirit with the perspective of science in practice.

To sum up, whether we consider a law as a tool for new discoveries or a guideline for manipulation, depends on whether we consider the law to represent the phenomenon we are interested in. This is not straightforwardly captured in the formulation of the theory, but depends on how the law is used in practice. So epistemic authority is not innately present in the laws. Even if we accept that the laws of thermodynamics express necessity, they only receive epistemic status in the practices where they are actually used to represent phenomena. It is worth mentioning that the importance of (contextual and pragmatic choices involved in) modelling practices is not commonly accepted among physicists:

> Thermodynamics is the much abused slave of many masters • physicists who love the totally impractical Carnot process, • mechanical engineers who design power stations and refrigerators, […] It is therefore natural that thermodynamics is prone to mutilation; different group-specific meta-thermodynamics' have emerged which serve the interest of the groups under most circumstances and leave out aspects that are not often needed in their fields. To stay with the metaphor of the abused slave we might say that in some fields his legs and an arm are cut off, because only one arm is needed; in other circumstances the brain of the slave has atrophied, because only his arms and legs are needed. Students love this reduction, because it enables them to avoid "nonessential" aspects of thermodynamics. But the practice is dangerous; it may backfire when a brain is needed. (Müller and Müller 2009, preface)

From the analysis presented here, I conclude that the necessity-approach to epistemic authority falls apart. Necessity, it seems, is not the way to understand why the Neuber rule is used and trusted in engineering practice. But then what is? In the following section, I will build on arguments by Sandra Mitchell to argue that engineers warrantedly use regularities like the Neuber rule, depending on the context. This will also enhance the alternative strategy to epistemic authority I mentioned in the introduction, namely the epistemic mark.

5.4 Contextual and pragmatic authority

Recall that traditionally the debate on laws consisted of two strategies to explain the epistemic authority laws get in science: laws express necessity, or laws have some epistemic mark. I spent the previous sections arguing that the first strategy does not explain why regularities like the Neuber rule are used in engineering sciences. In the current section, I will present an alternative that I believe does explain why the Neuber rule is used. This alternative expands the second strategy, namely the epistemic mark, and uses arguments by Sandra Mitchell, to make it more informative. I will first present Mitchell's contributions to the debate on laws in the life sciences and then adapt it to understand the Neuber rule and the engineering sciences more generally.

5.4.1 Mitchell's pragmatic account of laws

Sandra Mitchell developed her pragmatic account of laws in the context of biology. I already mentioned this matter in the introduction and in chapter 1. Let me briefly recapitulate. From about the 1970s to well in the 2000s, philosophers debated the nature of biological regularities, and more specifically, whether they should be considered laws. Beatty, for example argues against calling them laws, based on their contingency:

> […] all distinctively biological generalizations describe evolutionarily contingent states of nature— moreover, ''highly'' contingent states of nature in a sense that I will explain. This means that there are no laws of biology. For, whatever ''laws'' are, they are supposed to be more than just contingently true. (Beatty 1995, p.46)

In 1997, Mitchell distinguished 3 strategies of characterising laws: a normative, a paradigmatic and a pragmatic. The *normative* strategy encompasses approaches that start with a "definition of lawfulness" and then compare all candidate laws to this definition. If the specified conditions are met, the candidate qualifies as a law (Mitchell 1997, p.S469). Most of the accounts mentioned in section 5.1 are normative. Beatty's account is also a normative one. His definition includes natural necessity: laws are only "those generalisations that could never […] fail[ed] to be true" (Mitchell 1997, p.S469). This corresponds to the traditional debate focusing on necessity which I sketched above and which Cartwright and Frisch criticise. The second strategy, the *paradigmatic*, "begins with a set of exemplars of laws (characteristically in physics) and compares these to generalisations in biology" (Mitchell 1997, p.S469). I will not pay much attention to this strategy.

The final strategy, *pragmatism*, is the one Mitchell puts forward in her article (and goes on to develop throughout her later work). This is the one I will use to expand the epistemic mark account of laws. In Mitchell's pragmatic view, reference to definitions and exemplars is replaced with "an account of use of scientific laws" (Mitchell 1997, p.S475). According to Mitchell, we should entirely abandon the "received view" of what is required to be a law, viz.

> 1. logical contingency (having empirical content)
> 2. universality (covering all space and time)
> 3. truth (being exceptionless)
> 4. natural necessity (not being accidental)
>
> (Mitchell 2002b, p.330)

Instead, we should focus on how the generalisations in science are used. The specific contexts in and purposes for which generalisations are used can differ, naturally. Mitchell presents different parameters in virtue of which generalisations can be "evaluated for their usefulness":

> – Degree of accuracy attuned to specified goals of intervention
> – Level of ontology (populations vs individuals)
> – Simplicity: we use generalizations ranging from rules of thumb like Ptolemaic astronomical "laws" to navigate, to ideal gas laws that yield approximations within engineering tolerances.
> – Cognitive manageability: prior to the development of high-speed computation, mathematical equations were restricted to solvable linear formulations.
>
> (Mitchell 1997, p.S477)

The main point I want to take away from this for the current purposes is that, depending on the phenomenon we wish to study and the specific epistemic activity we are undertaking, different generalisations can prove more useful.[13]

Mitchell's project fits well with the focus of this chapter. Though Mitchell's account was developed for biology, I can use her framework to expand the analysis from the previous section. By building on her insights, we can get a better understanding of why

[13] I am not engaging with metaphysical questions, as this falls outside the scope of my thesis. But for those interested in a metaphysical companion story to my analysis, I recommend the work of Barry Ward (for instance (2002)).

the Neuber rule is successfully and warrantedly used in engineering sciences. This is the topic of the next section.

5.4.2 Pragmatic laws and epistemic authority

I believe that a pragmatic way of valuing regularities helps us understand the diversity of regularities that are used in the engineering sciences better than the necessity approach I discussed in section 5.3. On a pragmatic approach to laws, their epistemic authority does not result from any metaphysical necessity they express, but depends on the way they are used to model phenomena. Accordingly, different laws can gain more or less authority, depending on how successful they are with respect to the specific demands of the context.

As it is formulated, merely stating that epistemic authority of laws depends on the context does not give a more informative account of epistemic authority than the epistemic mark mentioned in section 5.1.2. Yet this is where Mitchell's account comes in. She has formulated several parameters by which we can understand and compare when regularities are best fit for the context and purposes at hand. Reflecting on different epistemic activities and goals that can be part of the engineering sciences provides a way of understanding why regularities like the Neuber rule are used in some contexts, and not in others.

In her original article, Mitchell distinguishes accuracy, ontology, simplicity and cognitive manageability as possible factors that influence the choice of regularity. Yet this is not an exhaustive list. In light of failure analysis specifically, I want to stress *feasibility* and *intelligibility* as important factors. In Carter's case of the collapsed spray drier, he needs to specify recommendations to modify the newer and stronger driers before they collapse as well. Because of this context, he is confronted with time-limitations and restrictions regarding redesign options. Given the task he faces, it is not useful to come up with a completely new design for a spray drier, since this will not influence the faith of the existing driers. So modelling the spray drier in terms of the materials with which it was constructed e.g., will not be ideal. Moreover, Carter needs to move fast and cannot spend months modelling the collapse of the drier in terms of more fundamental or micro laws and calculating all the variables. So the regularities he uses need to be intelligible.

Yet looking at the diversity of the engineering sciences, it's important to see that these demands differ when we consider different domains of the discipline. Note that this fits well with my focus on the user and context, inspired by Frisch. A scientist who wants to create a new material, more resistant to creep than others, may need to model

creep in more fundamental or micro terms. To understand the gravity of choices regarding modelling, consider another creep model. Mishin et al. (2013) developed a "general and rigorous theory of creep deformation". In their view, such a theory should contain

> (i) a thermodynamic model of a mechanically stressed crystalline solid with nonconserved lattice sites, (ii) a model of microstructure evolution that includes redistribution of vacancy sinks and sources and the motion of interfaces separating different phases and/or grains, and (iii) a set of kinetic equations derived from the entropy production rate and identification of the appropriate set of fluxes (including the creep deformation rate) and the conjugate driving forces. (Mishin et al. 2013, p.1)

They arrive at a sort of master equation, which they combine with assumptions about the physical properties of materials (e.g. whether it is isotopic, whether thermodiffusion cross-effects can be neglected) to derive a set of "phenomenological relations between fluxes and forces" that are part of this equation (Mishin et al. 2013, p.12). They apply this to an example and arrive at three equations that, with appropriate initial and boundary conditions, describe the entire dynamics of their system in deformed configuration. The equations are:

$$\frac{\partial \varphi}{\partial t} + v_L \nabla_x \varphi = -\frac{B}{T}\left[w'(\varphi) - \epsilon \nabla_x^2 \varphi\right]$$

$$\frac{\partial c_v}{\partial t} + v_L \nabla_x c_v - D_v \nabla_x^2 c_v = \nabla_x v_L$$

$$\nabla_x v_L = -B_r w(\varphi)\left[\frac{kT}{\Omega_0}\ln\frac{c_v}{c_v^0} - \sigma_{11}^{\infty} + w(\varphi) - \frac{1}{2}\epsilon(\nabla_x \varphi)^2\right]$$

In these equations c_v is the vacancy site fraction and $D_v = {k\Omega_0 L}/{c_v}$ the vacancy diffusion coefficient assumed to be constant, B_r a constant, σ_{11}^{∞} the coordinate-independent normal stress inside the grains, c_v^0 the equilibrium vacancy concentration in the absence of normal stress, $\Omega_0$the stress-free value of the volume per site (Ω), k Boltzmann's factor, $w(\varphi)$ a double-well function with an amplitude W creating a free-energy barrier between two lattice orientations, and ϵ is the gradient energy coefficient. These are the equations for a one-dimensional model. As they state,

> Due to the simplified geometry of this example, we will obviously not be able to model a real three-dimensional creep process taking place in polycrystalline materials. (Mishin et al. 2013, pp.15-16)

While this model has the potential to provide insight in creep in specific materials and can aid in explaining why certain materials behave the way they do, they will not likely

be helpful for a failure analyst like Carter[14]. But they might be useful in other epistemic contexts e.g. developing new materials. A pragmatic view on laws thus gives us a positive reason for engineers to use the Neuber rule: in certain contexts, the Neuber rule best fits the demands of the engineer and discipline.

The great diversity in approaches of creep-research seems to reflect this need for distinct regularities depending on the context. In some cases, the need for diversity is even explicitly acknowledged by engineering scientists. For example, Härkegård and Sørbø (1998) investigate the applicability of the Neuber rule because, regardless of the existing FEM techniques[15] to calculate stresses,

> […] it is still important for design engineers to have a qualitative notion of the key factors effecting stress and strain at notches. […] Therefore, validated and well-documented simplified methods for the approximate analysis of notches may still prove valuable. (p.224)

Correspondingly, the specific regularities that are used and trusted in various contexts will differ depending on the goal of the context and users. So whether and why a regularity receives epistemic authority is a question that can only be answered from within a specific context. This alternative, pragmatic view on laws thus presents a way to understand why engineers keep on using regularities like the Neuber rule. And while it builds on Goodman and others in the sense that the epistemic status of a regularity in a community is central for epistemic authority, the different parameters of usefulness help us get a better understanding of why certain regularities are trusted for certain purposes. In a sense, it is as Goodman said: laws are laws because they receive epistemic authority. This has been criticised as an anthropocentric and subjective criterion. And compared to the necessitarian view, it is. We could *in principle* have given epistemic authority to other regularities. But by formulating the parameters to evaluate whether a regularity is best fit, there is a less subjective way of giving regularities authority. At the very least, it is a mind-independent criterion. Moreover, the regularities also have to be based on evidence (which I have not spent time on here, but see chapter 4) and they need to be successful, they need to work. These criteria are all mind-independent. Moreover, which regularities receive epistemic authority in which contexts constantly

[14] Note that because this model cannot be used to describe three-dimensional phenomena, it also does not fulfil Nabarro's criteria mentioned in 4.1.

[15] FEM stands for Finite Element Methods, and refers to discretisation techniques in structural mechanics developed to solve mathematical equations by dividing them into non-overlapping components of simple geometry (Lin 2010, p.1).

changes. New regularities and applications are being developed (see above, the developments regarding creep) and they are being used and trusted. Understanding how and why this happens can best be done with a pragmatic story.

A pragmatic approach to epistemic authority also helps to bypass the theory-focus of traditional philosophy of science that I described in chapter 1 and that for instance Boon (2011) criticises. If the laws of physics (and equivalent laws) were the only ones that we can trust to make predictions and warranted explanations, then the engineers who rely on regularities like the Neuber rule would behave quite unsystematically and unmethodically. After all, their actions would not be guided by anything reliable, but the resulting diagnoses and predictions are trusted to make changes to existing artefacts, or to design new ones, as was clear from chapter 4. Because of all the successful applications and the general merits of engineering (sciences), this is highly unlikely and somewhat undervalues the methodology of a profession with great influence on our daily lives. A pragmatic understanding of laws, in combination with arguments against foundationalism and a focus on modelling, allows for a proper validation of the engineering sciences and their scientific practice.

The context-dependence of epistemic authority also helps to understand the distinction Boon and Knuuttila (2009) draw between engineering and the engineering sciences. They are different epistemic practices, with different goals and therefore different regularities. Looking back at the design recommendation Carter formulated for the spray drier (viz. to remove the lagging and cladding), it should be noted that this seems more straightforward than it might be. Actually designing a spray drier without the lagging and cladding might need some other adaptations in order for the resulting artefact to function in a stable way. Implementing the changes suggested in the design recommendations from failure analysts is a different epistemic practice than discovering what caused the failure. The first, I would say, is part of engineering design – a discipline with its own challenges and goals (see e.g. (Kroes 2009) and (Radder 2009)). The second is part of engineering sciences (since it aims at general knowledge). Correspondingly, the two practices might require different regularities to achieve their goals. Going from the recommendations to a new functioning artefact may require other regularities than the failure analysis, for instance regularities at another level, of a different specificity, knowledge of specific materials and threshold values, … This difference can also be

explained in the pragmatic approach to lawhood and epistemic authority I presented here.[16]

Finally, I want to reflect on the three strategies of defining laws that Mitchell distinguished. When we adopt a pragmatic view on regularities, this does not entail that in some contexts laws cannot be used normatively or paradigmatically by scientists. On the contrary, depending on the context, scientists can use the concept of law in a normative or paradigmatic way, because this is what the context of use requires. All of this is possible in a pragmatic account on laws and epistemic authority, while helping us understand why those regularities are used in that specific way. Because of this and all the other reasons above, I believe a pragmatic view on laws is remarkably well fit to reflect on scientific practice and specifically epistemic authority. If nothing else, it is better than the necessity approach. But hopefully, my analysis has shown that the pragmatic view can do more: it draws our attention to underinvestigated problems in philosophy of science and helps us understand the scientific practice of less visible disciplines such as the engineering sciences.

Conclusion

In this chapter, I investigated how we can legitimate why certain regularities receive *epistemic authority* in certain scientific practices. With "epistemic authority" I referred to the fact that regularities are trusted to achieve epistemic goals like prediction, explanation and manipulation. I tackled this question from the point of view of the engineering sciences, specifically failure analysis and used the Neuber rule from creep modelling as an exemplar.

I showed that in the philosophical literature, epistemic authority is often connected to the distinction between laws and mere regularities: laws can be trusted for epistemic goals, mere regularities cannot. Yet what makes a law a law is not agreed upon by philosophers. I discussed two common strategies for defining laws and for legitimating their epistemic authority: a necessitarian approach and an epistemic mark approach. Throughout the chapter I argued that neither was, in its current form, sufficient to

[16] For a detailed and informative discussion of the technology-engineering-science relation that corresponds with my analysis, see e.g. (Radder 2009), (Boon 2006) and (Boon 2011).

explain why the Neuber rule is trusted in engineering practice. Regarding necessity, I argued that the most obvious way to claim that the Neuber rule expresses necessity, was to derive it from more fundamental laws that are thought to express necessity. Building on Frisch's work in philosophy of physics, I then showed that (1) the Neuber rule is currently not successfully derived from (more) fundamental laws, that (2) the idea that there are truly fundamental laws that can be used to represent any phenomenon is not unproblematic given the functioning of scientific practice, and that (3) even if there are such fundamental laws, there is no guarantee that their necessity is undamaged by the modelling practices of science. I concluded that a necessitarian approach to epistemic authority does not help us to understand why the Neuber rule is trusted and used successfully in failure analysis.

As an alternative, I presented a pragmatic approach to epistemic authority, based on the work of Mitchell regarding laws of biology. I argued that whether a regularity receives epistemic authority depends on the specific demands and purposes of the scientific practice and undertaking. This entails that, even if we succeed in expressing the Neuber rule in more fundamental or micro terms, the resulting regularity might not receive epistemic authority in failure analysis, since it might be less apt to reach the specific goals of the discipline. I stressed that *feasibility* and *intelligibility* are important features for regularities in failure analysis. I argued that this pragmatic approach can explain the epistemic authority of the Neuber rule in failure analysis better than a necessitarian approach, while also accounting for the great diversity of regularities in different scientific disciplines. Moreover, I argued that this alternative account is more informative than the epistemic mark account of for instance Goodman. In this way, this chapter constituted an argument for my final specific claim (VI).

For my analysis, I also combined and expanded on arguments from philosophers of physics and philosophers of biology. As I explained in chapter 1, physics is often still seen as the exemplar, as the most mature science. The debate on laws in biology started from a comparison with laws of physics. Precisely because of the prestige that is connected with various sciences, with laws and with fundamentality, it is important that we combine insights from different fields in the philosophy of science. Thanks to philosophers like Frisch, who provide us with a more nuanced and practice-engaged view of physics, we can redraw the comparison. And this has consequences for other scientific disciplines as well. By moving away from a theory-focus view of physics as point of comparison and example for other sciences, the field opens up for legitimate research into different domains, like the engineering sciences.

As should be clear from the previous chapters in this dissertation, a whole range of philosophical debates are influenced by the definition and conception of law. As is clear from the way I conducted the analysis, the philosophical tools have long been in the

making. They are here, but need to be combined. I believe it is time we take the image that arises from combining them seriously: scientific practice does not differ as much across the domains as is often thought, and the way in which it differs is worth investigation. To give but one example: the relationship between philosophy of more traditional sciences (such as physics) and philosophy of engineering and technology. Similar points to Frisch's against foundationalism and for the importance of models have been made from the perspective of philosophy of technology. I already mentioned Boon's arguments. But Radder (2009) made a similar point in discussing the relation between science and technology: for fundamental theories to become empirically applicable, they have to be "developed and specified with a view to particular domains of empirical phenomena" (2009, p.72). He also defends the importance of modelling in science. Yet Frisch's points are still considered controversial, and philosophy of engineering and technology is still not booming in the way that philosophy of biology e.g. is. Integrating work from different debates can help strengthen the legitimacy of all these not so traditional disciplines. And that can, in my opinion, really benefit our understanding of science in all its forms and applications.

This chapter also presented the final complicating factor in my expose to show that using causal knowledge is a complex scientific practice, with its own problems, requiring philosophical attention (cfr. Generic aims (A) and (B)). Not only do we need to supplement information about regularities with excessive and detailed information about the physical setup or artefact and about the mechanism underlying the phenomenon, the regularities that we need are not exclusively the laws of physics. Because of all this, it is hard to maintain the view that we can understand all the aspects of physical causation by studying the laws of physics.

References

Andersen, Holly. 2017. What would Hume say? Regularities, laws, and mechanisms. In *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, edited by S. Glennan and P. Illari. London: Taylor & Francis.

Armstrong, David M. 1985. *What is a Law of Nature?*: Cambridge University Press.

Ashby, Michael F., and D. R. H. Jones. 2012. *Engineering Materials 1: An Introduction to Properties, Applications and Design*. 4 edition ed. Amsterdam: Butterworth-Heinemann.

Ayer, Alfred J. 1963. What is a Law of Nature? In *The Concept of a Person*: Springer.

Beatty, John. 1995. The evolutionary contingency thesis. In *Concepts, theories, and rationality in the biological sciences*, edited by G. Wolters and J. G. Lennox. Pittsburgh, Pa: University of Pittsburgh Press.

Boon, Mieke. 2006. How Science Is Applied in Technology. *International Studies in the Philosophy of Science* 20 (1):27-47.

———. 2011. In Defense of Engineering Sciences: On the Epistemological Relations Between Science and Technology. *Techne* 15 (1):49-71.

Boon, Mieke, and Tarja Knuuttila. 2009. Models as epistemic tools in engineering sciences: a pragmatic approach. *International Journal of Software Engineering and Knowledge Engineering*:687-720.

Braithwaite, Richard Bevan. 1953. *Scientific explanation: A study of the function of theory, probability and law in science*: CUP Archive.

Brandon, Robert N. 1997. Does biology have laws? The experimental evidence. *Philosophy of Science*:S444-S457.

Brown, L. M. 2010. Frank Reginald Nunes Nabarro MBE. 7 March 1916 — 20 July 2006. *Biographical Memoirs of Fellows of the Royal Society* 56:273-283.

Carrier, Martin. 1995. Evolutionary change and lawlikeness: Beatty on biological generalizations. In *Concepts, theories and rationality in the biological sciences*, edited by G. Wolters and J. G. Lennox.

Carroll, John W. 1994. *Laws of nature*: Cambridge University Press.

———. 2016. Laws of Nature. In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta: Metaphysics Research Lab, Stanford University.

Carter, Paul. 2001. Creep Failure of a Spray Drier. In *Failure Analysis Case Studies II*, edited by D. R. H. Jones. Amsterdam: Pergamon.

Cartwright, Nancy. 1983. *How the laws of physics lie*: Oxford University Press.

Craver, Carl F., and Marie I. Kaiser. 2013. Mechanisms and Laws: Clarifying the Debate. In *Mechanism and Causality in Biology and Economics*, edited by H.-K. Chao, S.-T. Chen and R. L. Millstein: Springer Netherlands.

Dretske, Fred I. 1977. Laws of Nature. *Philosophy of Science* 44 (2):248-268.

Failure Analysis Case Studies II. 2001. Edited by D. R. H. Jones. 1 edition ed. Amsterdam: Pergamon.

Frisch, Mathias. 2014. *Causal reasoning in physics*: Cambridge University Press.

Goodman, Nelson. 1983. *Fact, Fiction, and Forecast, Fourth Edition*. 4th Revised ed. edition ed. Cambridge, Mass: Harvard University Press.

Gould, Stephen Jay. 1990. *Wonderful Life: The Burgess Shale and the Nature of History*: W. W. Norton.

Härkegård, G., and S. Sørbø. 1998. Applicability of Neuber's rule to the analysis of stress and strain concentration under creep conditions. *Transactions- American Society of Mechanical Engineers Journal and Technology* 120:224-229.

Kroes, Peter. 2009. Foundational Issues of Engineering Design. In *Philosophy of Technology and Engineering Sciences*, edited by A. W. M. Meijers, D. M. Gabbay, P. Thagard and J. Woods. Amsterdam: North Holland.

Kuhlmann, Meinard. 2017. Mechanisms in physics. In *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, edited by S. Glennan and P. Illari. London: Taylor & Francis.

Lange, Marc. 2000. *Natural laws in scientific practice*: Oxford University Press.

Lewis, David K. 1973. *Counterfactuals*. Cambridge: Harvard University Press.

Lin, Liwei. 2010. Introduction to Finite Element Modeling. Teaching document for ME 128 Computer-Aided Mechanical Design, Berkeley.
Mill, John Stuart. 1911. *A System of Logic. Eighth, reprinted ed*: London: Longmans, Green and Co.
Mishin, Y., J. A. Warren, R. F. Sekerka, and W. J. Boettinger. 2013. Irreversible thermodynamics of creep in crystalline solids. *Physical Review B* 88 (18).
Mitchell, Brian S. 2004. *An Introduction to Materials Engineering and Science for Chemical and Materials Engineers*: John Wiley & Sons.
Mitchell, Sandra D. 1997. Pragmatic Laws. *Philosophy of Science* 64:S468-S479.
———. 2002a. Ceteris paribus—an inadequate representation for biological contingency. *Erkenntnis* 57 (3):329-350.
———. 2002b. Contingent generalizations: Lessons from biology. *Akteure–Mechanismen–Modelle: Zur Theoriefähigkeit makro-sozialer Analysen*:179-95.
Müller, Ingo, and Wolfgang H. Müller. 2009. *Fundamentals of Thermodynamics and Applications: With Historical Annotations and Many Citations from Avogadro to Zermelo*: Springer Science & Business Media.
Nabarro, Frank R. N. 2004. Do we have an acceptable model of power-law creep? *Materials Science and Engineering: A* 387–389:659-664.
Norton, John D. 2003. A Material Theory of Induction. *Philosophy of Science* 70 (4):647-670.
Psillos, Stathis. 2002. *Causation and explanation*. Vol. 8: McGill-Queen's Press-MQUP.
Radder, Hans. 2009. Science, Technology and the Science–Technology Relationship. In *Philosophy of Technology and Engineering Sciences*, edited by A. W. M. Meijers, D. M. Gabbay, P. Thagard and J. Woods. Amsterdam: North Holland.
Ramsey, Frank. 1928. Universals of Law and of Fact.
Readings on laws of nature. 2004. Edited by J. W. Carroll. Pittsburgh, Pa: University of Pittsburgh Press.
Tooley, Michael. 1977. The nature of laws. *Canadian Journal of Philosophy* 7 (4):667-698.
Van Fraassen, Bas C. 1989. *Laws and Symmetry*: Oxford University Press.
Ward, Barry. 2002. Humeanism without Humean Supervenience: A Projectivist Account of Laws and Possibilities. *Philosophical Studies* 107 (3):191-218.
Weber, Erik, Jeroen Van Bouwel, and Leen De Vreese. 2013. *Scientific Explanation*: Springer Science & Business Media.

Conclusion

In this dissertation I have investigated how we successfully use and produce appropriate physical causal knowledge to design, explain, maintain and repair artefacts. Causation in physics is traditionally discussed from a very theoretical point of view, focusing on the fundamental laws of physics. In this work, I have shown that when we focus instead on how we develop and use physical causal knowledge to achieve epistemic goals (i.e. when we take a use-perspective), many complexities that are not captured by the theoretical discussion become tangible. In chapter 1, I sketched the framework in which this dissertation is to be understood. I showed that laws take up a central position in the philosophy of physics, and that this influenced the philosophy of other sciences as well. Recently, philosophy of the special sciences (viz. the biomedical and the social sciences) has let go this focus on laws. In similar spirit, the Society for the Philosophy of Science in Practice (SPSP) has argued that we need to study science as a practice, and not limit ourselves to the theory or laws. However, philosophy of physics remains very focused on laws and theory. To supplement this, I analysed cases where we use physical causal knowledge with regards to the meaning of this knowledge, the evidence for this knowledge and its relation to laws. The cases were of different complexity levels and addressed different sources of complexities.

Recall that in the introduction, I formulated 4 specific claims that I would argue for throughout the dissertation:

(I) To account for our successful creating, explaining, repairing, and maintaining of artefacts, we need a lot of specific physical causal information of the right kind, both of the artefact and of the physical and social context it functions in.

(II) The evidence needed to argue for physical causal claims extends beyond the laws of physics. We also use mechanistic evidence.

(III) The laws of physics are not the only source of general knowledge that is used to reach epistemic goals. Another important source is generalising local causal knowledge

(IV) When looking at use, the importance of the distinction between laws and non-laws becomes significantly less prominent and the focus shifts to contextual goals. As such, the focus that philosophy of physics puts on theory and fundamental laws does not aid us in understanding how we use physical causal knowledge to achieve epistemic goals.

In chapters 2 till 5, I argued for these claims. In chapter 2, I argued for the first claim. I studied causal knowledge that needs to underlie remedy instructions from bike, car and radio repair manuals. This chapter showed that even for day-to-day artefacts in rather easy contexts, we still need a lot of information in order to warrant causal claims that we can use to intervene in the world. It also showed that the validity and meaning of many physical causal claims depend on the context, both on the physical and the social. I furthermore argued that depending on the demands we put on our interventions, we require different properties of the causal relation. For the manuals, this included positive causal factorhood, weak context-unanimity and Mackie causation. This chapter dealt with the least complex cases of my dissertation, since they involved reliable knowledge that was very focused on the interventions.

In chapter 3, I turned my attention to claim to and focused on evidence. I argued that in reality, we need to *support* causal claims. I discussed how we support causal claims occurring in causal explanations. The laws of physics do not suffice to support them. Physical laws only provide what I called correlational evidence. I argued that for phenomena that fall under the laws of physics, we need mechanistic evidence to supplement the evidence from physical laws. The interplay between the laws of physics and mechanistic evidence resembles the interplay between correlations in the special sciences and mechanistic evidence. I showed that the main difference is related to the mathematical equivalency that most laws express, which often makes it easier to exclude the possibility of common causes and nonsense correlations in physical cases, than it is in the biomedical or the social sciences. The chapter showed that supporting causal claims that we use in explanations, also requires a lot of information of the right kind. And, more importantly, this evidence is not just laws, we also need information about mechanisms. This constituted the argument for claim (II).

In chapter 4, I investigated practices where the laws of physics do not function as source of general physical knowledge, but where causal knowledge from specific cases was generalised to other situations. This adds another layer of complexity to the cases compared to the ones from chapter 2 and 3. I discussed 3 examples from failure analysis.

I showed that to make the knowledge generalisations warranted, engineers need information about the mechanism of the artefact, both to determine the domain of the generalisation and to perform a comparative processing test. Because these cases deal with malfunction of rather complex machinery, the laws of physics are not the most straightforward source of general knowledge. I argued that this is the case for many contexts in which we want to use physical causal knowledge, and that finding this knowledge is not easy. In this way, I argued for claim (III).

In chapter 5, I turned to the last claim (viz. (IV)). In a sense, the previous chapters provided groundwork for this final chapter. In it, I reflected on a regularity (viz. the Neuber rule) from failure analysis that receives epistemic authority and I argued that traditional views that connect this authority to lawhood cannot help us understand why this regularity is trusted for interventions and explanations. I used arguments from Mathias Frisch to show that regularities only get epistemic authority for phenomena that are actually modelled with the regularities or are sufficiently close to such phenomena. Relatedly, I showed that the view that there are fundamental laws that can be used to represent any phenomenon is not unproblematic. And because models do not follow from regularities in necessary way, attempting to explain the authority of the Neuber rule via the necessity of the laws of physics is not a successful strategy. I argued that a pragmatic view on epistemic regularity fits better. Based on Sandra Mitchell's contributions to the philosophy of the life sciences, I presented different parameters according to which the usefulness of regularities can be determined, like feasibility and intelligibility. Correspondingly, the difference between laws and non-laws becomes less important than whether a regularities is suited for achieving the specific goals. In this way, I also refuted an assumption that is present in the philosophy of physics, viz. that the laws of physics are the main source of general knowledge we need to achieve our epistemic goals.

The combination of all these chapters and conclusions allow me to reach the two generic aims that I specified in the Introduction:

(A) To show that using and producing physical causal knowledge is not a trivial affair.

(B) To show there a several serious philosophical issues connected to useful physical causal knowledge.

As explained, these aims would be reached throughout the dissertation, rather than in one specific chapter. Now that all my arguments have been presented, it is clear that together, they fulfilled (A) and (B). All chapters identified reasons why using and producing physical causal knowledge is not trivial. For instance, we need to be certain

that the meaning of the knowledge is appropriate for our envisioned epistemic goals, we require a lot of evidence that is not easy to come by and determining which regularities are to be trusted for our specific situations is far from straightforward. By attempting to analyse and understand these difficulties, I have shown that, pace what philosophers like Norton assume, these difficulties are serious philosophical issues, worthy of reflection.

Let me take stock of what these conclusions mean. The first important point I want to stress is that the focus on laws does not help us understand how we explain, and intervene in, the world based on physical causal knowledge. From the way I have put contributions from different literatures together, it should be clear that the philosophical tools to support this point are there. Philosophy of the special sciences is turning to practice and use, philosophy of the engineering sciences is dedicated to a more practice-engaged part of science, and philosophy of technology provides a rich basis for analysing artefacts. Philosophers like Cartwright, Mitchell, Steel, Frisch and all the mechanists have been attempting to refocus philosophy of science more towards practice for some time. And yet, laws remain to be very central in philosophy of physics, and paying attention to physical applications, is still controversial. I have shown that this philosophical focus on laws rests on assumptions that are hard to maintain when we are prepared to focus on *using* physical causal knowledge.

I have not just advocated a turn to scientific practice of physics. I have advocated a turn to practices surrounding *all kinds of artefacts*, including common household ones. The development and functioning of common artefacts is often taken for granted and simultaneously seen as a result of science. I have shown that developing, explaining, maintaining and repairing artefacts, even common household ones, is a complex practice. It cannot be understood by studying a select group of physical laws that often have rather little to do with this practice. Rather, a combination of insights from philosophy of physics, philosophy of technology and philosophy of engineering is needed for understanding practices related to artefacts. My dissertation is an example that shows what can be gained from combining these literatures. I hope it encourages others to do the same, and in this way bring these domains closer together. A related consequence of my analysis concerns the focus that philosophy of physics lies on theoretical physics. Again, it goes without saying that this is an important domain. But there is so much more out there. Fundamental physics differs significantly from the practices that I described in this dissertation, like bike repair and engineering. It is very conceivable that we may need different philosophical tools to study the different practices. Even when we study theoretical physics from a PSP-perspective, it will not give us answers to all the philosophical questions raised here, questions related to how physical knowledge influences our day-to-day lives. So I also plead for a more liberated view of which domains of physical causal knowledge are philosophically reflected upon –

a view that takes the materialisations of physical knowledge seriously. The practices I discussed are real, and all but easy. They require philosophical attention. I hope to have provided some starting point for doing so.

Correspondingly, the picture that arises is one where physical practices are remarkably similar to practices from the special sciences. This is another point I want to emphasize. I used philosophical insights from the special sciences to elucidate the physical cases in every chapter. The reason why this methodology worked is because in many physical practices, the same problems or difficulties arise as in the special sciences. Think about the need for mechanistic evidence, the need for knowledge generalisation techniques, the contingency of regularities. There are of course differences between the physical practices I described and the special sciences, just like there are differences among the different special sciences. But from my analysis, it should be clear that those differences are significantly less prominent than is often assumed in philosophy of science. More importantly, the ways in which the practices differ is worth investigation. But all of these questions require that we relax the philosophical focus on the laws and theory of physics. In that sense, I hope that this dissertation can also contribute to the status and prestige of certain not so traditional disciplines. When we accept that epistemic practices related to the exemplar science (viz. physics) are in fact rather similar to practices in the special sciences, the ideal that sciences need to live up to changes, and correspondingly, the status of other (not so) scientific disciplines changes as well. And provided we believe that philosophy has some impact on society, this can have influences on funding, on how resources are spent and prioritised, on when scientific results are thought to be established.

Because of all my points, let me conclude by suggesting that it is the focus on laws in philosophy of physics that “is a relic of a bygone age”, and not causation, as was said to be in the beginning of this dissertation (see the introduction).

Summary

In this dissertation I have investigated how we successfully use physical causal knowledge to intervene in and explain the physical world. By doing so, I have attempted to bring a more practice engaged voice to philosophy of physics in general, and to the characterisation of causation in physics in specific. The question of whether the laws of physics express causal information has drawn philosophical attention for some time, with most philosophers answering negatively. By and large, their arguments are based on the symmetry that most laws of physics express. For one, laws of physics are often mentioned as mathematical equations. Hence, the value of every variable in the equation can be calculated from the values of the other variables. This stands in contrast with causal relations: the value of the cause can be used to determine the value of the effect, but often not vice versa. Moreover, causal relations are thought to be time-asymmetric: causes precede their effects. Laws of physics, on the other hand, are time-symmetric: when you know the state of a physical system at a certain point in time, you can determine both the past and future states of the system. Because of these dissimilarities, philosophers have concluded that there is no causal information in the laws of physics. This conclusion does not seem compatible with day-to-day practice, since we use physical causal information to intervene in the natural world and to explain it. Sending spacecrafts to the moon and ensuring that they – and the astronauts flying them – come back for example, requires the use of mechanics and other physical knowledge. These are goal-directed activities, with an inherent asymmetry between cause (means) and effect (goal). In this dissertation, I attempted to resolve this contrast by studying how we use physical causal knowledge and what this can teach us about physics and the laws of physics.

In chapter 1, I explained this proposed in shift in focus. It is clear that we use physical causal knowledge often, among other things for designing, building, maintaining, explaining and repairing artefacts. These are the contexts I focused on throughout this PhD. Such a focus on applications is notoriously absent from philosophy of physics, since

applications are thought to follow straightforwardly from the laws. I argued that this is connected with two assumptions: that the mathematical expression of the laws carry all their information and that the laws exhaust the content of physics.

In this dissertation, I built on existing philosophical analyses developed in light of the special sciences (viz. the social and biomedical sciences), to show that using physical causal knowledge is not straightforward at all, and does raise interesting philosophical issues that cannot be resolved by focusing on the laws of physics. Moreover, when we thoroughly look at applications, the claim that physics contains no causal information turns out to be untenable as well. I have argued for this by discussing different contexts of use with increasing complexity, ranging from technical repair manuals, over common artefacts to failure analysis. Each of these contexts allowed me to draw focus to a different interesting philosophical problem related to using and producing physical causal knowledge.

In chapter 2, I studied the meaning of causal claims on which we want to base interventions. For this chapter, I used examples from technical repair manuals for bikes, cars and radio's as examples. I showed that depending on the demands we put on the results of our interventions, we need different causal knowledge. Moreover, causal relations depend on the context. While in the case of a heated pressure cooker, the temperature determines the pressure, the causal relation is reversed in other artefacts. This chapter showed that even for day-to-day artefacts in rather easy contexts, we still need a lot of information in order to warrant causal claims that we can use to intervene in the world. This constituted the first important issue related to using and producing physical causal knowledge. So if we are interested in using causal knowledge, we need a lot of information that is not straightforwardly contained in the laws. This is why repair manuals are so useful. They prescribe actions that fulfil our goals without us needing to find all the information needed to warrant the actions ourselves. They have authority.

However, in many cases, there is no such manual and we need to support the causal claims we want to base interventions and explanations on ourselves. This was the topic of chapter 3, where I discussed explanations in the context of common artefacts. Based on those cases, I showed that supporting causal claims is not that straightforward either. Suppose we want to explain why a fire occurred. We may refer to the cause of the fire. For example, we might say that the shortcut caused the fire. For our explanation to be successful, we need to support this causal claim. I showed that the laws of physics do not suffice to support causal claims. They only provide what I called correlational evidence. I argued that for phenomena that fall under the laws of physics, we need mechanistic evidence to supplement the evidence from physical laws. This means we need information about the mechanism that connected the shortcut to the fire, and this

information is not given by the laws. This is the second important philosophical issue related to using and producing physical causal knowledge.

In chapter 4, I studied engineering practices. Because of the artefacts that engineers handle and the contexts they deal with, these cases are more complicated than the day-to-day contexts I discussed in previous chapters. I looked at failure analysis, an engineering practice where failures of artefacts are investigated to guide maintenance and design practices of other artefacts. Analysts generalise knowledge from one failure to other contexts. This is another source of general physical knowledge than the laws of physcs. I investigated how the engineers support their generalisations. I showed that to make the knowledge generalisations warranted, engineers need information about the mechanism of the artefact and about the broader context the artefact functions in. This information is not contained in the laws of physics either, so the way that this information is gathered and characterised, forms the third important issue related to using physical causal knowledge.

In chapter 5, I turned to the relation between practices surrounding artefacts and the laws of physics. I focused on why regularities receive what I called epistemic authority in the context of artefacts. With epistemic authority, I referred to the fact that certain information is trusted to base interventions or explanations on. Traditionally, this authority has been linked to lawhood: fundamental laws deserve this authority. In engineering practices, other regularities are often used to guide interventions or explanations. Yet these regularities are often considered to be less than the laws of physics: they are not universally valid and they do not express any necessity. I argued that the view that universal, necessary laws are more appropriate to achieve our epistemic goals is mistaken. In reality, a whole array of regularities is used to achieve goals, and which regularities are used depends on the context and the specific goals we have. Correspondingly, I showed that the laws of physics are not straightforwardly the main source of information for reaching epistemic goals. Deciding and understanding which regularities are used and why, constituted the final important philosophical issue related to using physical causal knowledge that I identified in this dissertation.

These arguments showed that the assumption that using and producing physical causal knowledge is straightforward and unproblematic, is mistaken. Correspondingly, the focus on laws in philosophy of physics is not unproblematic, and actually hinders philosophical reflection on the broader physical sciences. By looking at applications, the gap between physics and the special sciences turned out to be significantly smaller than often assumed. At the same time, it became clear that our understanding of science would benefit from an increased interplay between philosophy of physics on the one hand and philosophy of engineering and of technology on the other.

Correspondingly, when we accept that the exemplar science (viz. physics) has in fact more in common with the special sciences, the ideal that sciences need to live up to changes. This also influences the status of other scientific disciplines. And provided we believe that philosophy has some impact on society, a changed perception of physics can influence funding, how resources are spent and prioritised, and even change when scientific results are thought to be established. In general, a practice-based philosophy of physics can significantly alter how we characterise and value science as a whole.

Samenvatting

In dit doctoraat heb ik onderzocht hoe we op succesvolle wijze fysische causale kennis gebruiken om de fysische wereld te begrijpen, en deze te manipuleren. Met deze focus heb ik geprobeerd een meer praktijk geïnspireerde visie te introduceren in filosofie van de fysica en dan specifiek in het debat rond causaliteit in fysica. Filosofen buigen zich al lang over de vraag of de wetten van de fysica causale informatie bevatten. Volgens het merendeel van de hedendaagse filosofen van de fysica is het antwoord negatief. In het algemeen zijn hun argumenten gebaseerd op het feit dat de meeste wetten van de fysica symmetrisch zijn. Eerst en vooral worden de wetten vaak uitgedrukt als wiskundige vergelijkingen. Zodoende kan de waarde van elke variabele in de vergelijking berekend worden op basis van de waarden van de andere vermelde variabelen. Dit staat in contrast tot causale relaties: de waarde van de oorzaaksvariabele kan gebruikt worden om de waarde van de effectsvariabele te beïnvloeden, maar niet andersom. Bovendien wordt van causale relaties vaak gesteld dat ze een tijds-asymmetrie uitdrukken: oorzaken gaan vooraf aan effecten. De wetten van de fysica daarentegen, zijn tijds-symmetrisch: ze laten je toe om op basis van de toestand van een fysisch systeem op één bepaald tijdstip, zowel de toekomstige als voorbije toestanden van dat systeem te bepalen. Omwille van deze verschillen met betrekking tot symmetrie hebben veel filosofen geconcludeerd dat de wetten van de fysica geen causale informatie bezitten. Deze conclusie lijkt echter niet compatibel met dagdagelijkse praktijk, gezien we fysische causale informatie constant gebruiken om de natuurlijke wereld te verklaren en erin in te grijpen. Denk bijvoorbeeld maar aan ruimtevaart. Shuttles naar de ruimte sturen en ervoor zorgen dat ze veilig terugkeren, vraagt veel kennis van mechanica en andere fysische domeinen. Dit zijn doelgerichte activiteiten, met een cruciale asymmetrie tussen de oorzaak (het middel) en het gevolg (het doel). In dit doctoraat heb ik geprobeerd dit contrast tussen dagdagelijkse en wetenschappelijke praktijk aan de ene kant en filosofie van de fysica aan de andere kant, op te lossen. Dit heb ik gedaan door te focussen op

hoe we fysische causale kennis gebruiken en produceren, en wat dit ons kan leren over (de wetten van de) fysica.

In hoofdstuk 1 heb ik deze verandering van focus gekaderd. Uit een blik op de praktijk blijkt dat we fysische causale kennis onder meer gebruiken voor het ontwerpen, bouwen, onderhouden, verklaren en repareren van artefacten. Dit zijn de contexten waaraan ik aandacht besteed heb in dit doctoraat. Zo'n focus op toepassingen is notoir afwezig in filosofie van de fysica, omwille van de overtuiging dat toepassingen een direct en onproblematisch gevolg zijn van de wetten van de fysica. Ik heb geargumenteerd dat deze overtuiging verbonden is met twee vooronderstellingen, namelijk dat de wiskundige formuleringen van wetten de volledige inhoud van de wet weergeven en dat de wetten van de fysica de volledige inhoud van het wetenschappelijk domein vatten.

In dit doctoraat heb ik filosofische analyses uit de speciale wetenschappen (viz. de sociale en biomedische wetenschappen) aangepast en uitgebreid om aan te tonen dat fysische causale kennis produceren en gebruiken absoluut niet evident is en weldegelijk belangrijke filosofische problemen genereert. Bovendien kunnen deze problemen niet opgelost worden door louter de wetten van de fysica te bestuderen. Daarnaast heb ik aangetoond dat een grondige focus op het gebruik van fysische causale kennis niet compatibel is met de idee dat fysica geen causale informatie bevat. Ik heb deze punten beargumenteerd door verschillende contexten waarin fysische causale kennis gebruikt wordt te bestuderen. Deze contexten werden gradueel complexer, gaande van technische handleidingen over alledaagse artefacten tot de wetenschappelijke analyse van gefaalde artefacten. Elk van deze contexten gaf aanleiding tot een nieuw interessant filosofisch probleem gerelateerd aan het gebruiken en produceren van fysische causale kennis.

In hoofdstuk 2 heb ik de betekenis bestudeerd van causale beweringen waarop we interventies willen baseren. Voor dit hoofdstuk heb ik voorbeelden uit technische reparatie handleidingen van fietsen, auto's en radio's gebruikt als casestudies. Ik heb aangetoond dat het type causale kennis dat we nodig hebben voor onze interventies, afhangt van de specifieke eisen die we stellen aan het resultaat van die interventies. Bovendien zijn causale relaties context-afhankelijk. In een verhitte snelkookpan bepaalt de temperatuur de druk in de pan, maar voor andere artefacten is de causale relatie omgekeerd. Dit hoofdstuk toonde aan dat we zelfs voor zeer alledaagse artefacten in vrij simpele omstandigheden, veel informatie nodig hebben om causale beweringen waarop we interventies willen baseren te rechtvaardigen. Dit vormde het eerste belangrijke filosofische probleem met betrekking tot het gebruik van fysische causale kennis. De informatie die nodig is om fysische causale kennis te gebruiken op een nuttige manier zit niet overduidelijk in de wetten. Dit is waarom technische reparatie handleidingen zo nuttig zijn. Ze specifiëren handelingen die onze doelen vervullen zonder dat we zelf de

informatie moeten verzamelen die nodig is om de handelingen te rechtvaardigen. De handleidingen hebben autoriteit.

In veel situaties zijn er echter geen handleidingen beschikbaar en moeten we zelf verantwoording geven voor de causale beweringen waarop we interventies en verklaringen willen baseren. Dit heb ik besproken in hoofdstuk 3. Daar heb ik aangetoond dat causale beweringen verantwoorden in de context van verklaringen ook niet evident is. Wanneer we bijvoorbeeld willen verklaren waarom er ergens brand uitbrak, kunnen we verwijzen naar de oorzaak van de brand. We kunnen bijvoorbeeld stellen dat er een kortsluiting had plaatsgevonden. Opdat onze verklaring succesvol zou zijn, moeten we deze bewering staven. Ik heb aangetoond dat de wetten van de fysica niet volstaan om zulke fysische causale beweringen te staven. We hebben ook informatie nodig over het mechanisme dat de kortsluiting met de brand verbond. Dit was het tweede belangrijke filosofische punt dat ik identificeerde met betrekking tot het gebruik van fysische causale kennis.

In het vierde hoofdstuk heb ik ingenieurswetenschappen bestudeerd, en specifiek de analyse van gefaalde artefacten. Het soort artefacten waar ingenieurs mee in contact komen, en de contexten waarin de ingenieurs werken, zijn complexer dan de meer alledaagse contexten en artefacten uit de vorige hoofdstukken. Ik heb mij specifiek gefocust op ingenieurspraktijken waar gefaalde artefacten bestudeerd worden met als doel ontwerp- en onderhoudspraktijken van andere artefacten te verbeteren. Gespecialiseerde ingenieurs generaliseren kennis van één falen naar andere contexten. Dit is een andere bron van algemene fysische kennis dan de wetten van de fysica. Ik heb onderzocht hoe ingenieurs hun generalisaties verantwoorden. Ik heb aangetoond dat ze hiervoor informatie over het mechanisme van het artefact nodig hebben, alsook informatie over de bredere context waarin het artefact functioneert. Deze informatie zit ook niet rechtstreeks in de wetten van de fysica. De manier waarop deze informatie verzameld wordt en hoe ze te karakteriseren vormde het derde belangrijke filosofische probleem dat ik geïdentificeerd heb in verband met fysische causale kennis gebruiken en produceren.

Hoofdstuk 5 handelde over de relatie tussen praktijken rond artefacten en de wetten van de fysica. Ik heb onderzocht waarom regulariteiten epistemische autoriteit ontvangen in de context van artefacten. Met epistemische autoriteit bedoelde ik het feit dat bepaalde informatie vertrouwd wordt om interventies en verklaringen op te baseren. Traditioneel werd deze autoriteit verbonden met wetmatigheid: fundamentele wetten verdienen de autoriteit. In ingenieurspraktijken worden echter andere regulariteiten dan fundamentele wetten gebruikt om interventies en verklaringen op te baseren. Deze regulariteiten worden vaak als minder informatief of betrouwbaar beschouwd dan de wetten van de fysica: de regulariteiten zijn niet universeel geldig en ze drukken geen

noodzakelijk verband uit. Ik heb geargumenteerd dat de idee dat universele, noodzakelijke wetten beter zijn om onze epistemische doelen te bereiken, onhoudbaar is. In realiteit wordt een hele waaier van regulariteiten gebruikt, afhankelijk van de context en de specifieke doelen die we willen bereiken. Overeenkomstig heb ik aangetoond dat de wetten van de fysica niet eenduidig de belangrijkste bron van informatie zijn die we nodig hebben voor interventies en verklaringen. Beslissen en begrijpen welke regulariteiten wanneer gebruikt worden en waarom, vormde het laatste belangrijke filosofische probleem dat ik geïdentificeerd heb in dit doctoraat.

De combinatie van deze hoofdstukken en argumenten toont de onhoudbaarheid aan van de veronderstelling dat het gebruiken en produceren van fysische causale kennis evident is. Overeenkomstig is de eenzijdige focus op wetten in filosofie van de fysica ook niet onproblematisch en hindert deze focus filosofische reflectie op bredere fysische praktijk (viz. onder meer praktijken rond artefacten). Door de focus te verleggen naar toepassingen bleek dat de speciale wetenschappen en fysica meer gemeen hebben dan vaak wordt aangenomen in wetenschapsfilosofie. Tegelijkertijd werd duidelijk dat ons begrip van wetenschappelijke praktijk baat zou hebben bij een grotere kruisbestuiving tussen filosofie van de fysica en filosofie van de ingenieurswetenschappen en van de technologie.

Wanneer we bovendien inzien dat de voorbeeldwetenschap bij uitstek (namelijk fysica) meer gemeen heeft met de speciale wetenschappen, dan wijzigt ook het ideaalbeeld dat we hebben van wetenschap. Deze wijziging heeft op zijn beurt gevolgen voor het wetenschappelijk statuut van bepaalde disciplines zoals bijvoorbeeld technisch ontwerp. En mits we geloven dat filosofie enige impact heeft op de samenleving, kan deze wijziging in perceptie van wat wetenschap is invloed hebben op fondsenverstrekking, op hoe budgetten gespendeerd worden en zelfs op wat telt als een vaststaand wetenschappelijk resultaat. Concluderend kan gesteld worden dat een meer praktijk- en toepassingsgerichte filosofie van fysica de maatschappelijke en filosofische visie op en valuatie van wetenschap significant kan beïnvloeden.